

TEKNILLINEN KORKEAKOULU
Tietotekniikan osasto

Juha Aatrokoski

QDS: ohjelmisto pistelähteiden erotteluun Planck-satelliitin datasta



Diplomityö

Espoo, 30. lokakuuta 2006

Valvoja: Professori Eljas Soisalon-Soininen
Ohjaaja: Anne Lähteenmäki, TkT

HELSINKI UNIVERSITY OF TECHNOLOGY Department of Computer Science and Engineering		ABSTRACT OF MASTER'S THESIS	
Author Juha Aatrokoski			
Date 30th October 2006		Pages 58	
Title of thesis QDS: Software For Point Source Extraction From Planck Satellite Data			
Professorship Software Technology		Professorship Code T-106	
Supervisor Professor Eljas Soisalon-Soininen			
Instructor Anne Lähteenmäki, D.Sc. (Tech.)			
<p>Planck is a satellite of the European Space Agency ESA that explores the cosmic background radiation and is to be launched in 2008. In addition to the background radiation the Planck project also researches other components of the radio sky, such as active galactic nuclei and the Milky Way galaxy.</p> <p>The radiation of active galactic nuclei can vary significantly even within a few days. As the Planck results don't become available until the end of the mission, a software is needed in the research of active galactic nuclei that separates and analyses point sources from the Planck data in near real time so that results can be gotten after only a few days from the actual observation. Based on the results of the software other observation facilities are alerted for more accurate follow-up observations. The subject of this master's thesis is the final implementation of that software.</p> <p>The software processes data from the Planck low frequency instrument on observation frequencies 30, 44 and 70 GHz. The data is processed in one hour periods that each form a ring in the sky. The noise in the signal is reduced and point sources are separated by filtering the data with a filter determined by the Mexican hat wavelet. The point sources found are compared to the previous observations stored in a database and an alert is produced for those that are deemed interesting. The final decision about follow-up observations is made by the software operator. The software will function in the data processing centre of the Planck low frequency instrument in Trieste, Italy, throughout the entire operation of the satellite.</p>			
Keywords radio astronomy, signal analysis, Mexican hat wavelet, active galactic nuclei			

TEKNILLINEN KORKEAKOULU		DIPLOMITYÖN TIIVISTELMÄ	
Tietotekniikan osasto			
Tekijä			
Juha Aatrokoski			
Päiväys		Sivumäärä	
30. lokakuuta 2006		58	
Työn nimi			
QDS: ohjelmisto pistelähteiden erotteluun Planck-satelliitin datasta			
Professuuri		Koodi	
Ohjelmistotekniikka		T-106	
Työn valvoja			
Professori Eljas Soisalon-Soininen			
Työn ohjaaja			
Anne Lähteenmäki, TkT			
<p>Planck on Euroopan avaruusjärjestö ESAn vuonna 2008 laukaistava kosmista taustasäteilyä tutkiva satelliitti. Taustasäteilyn ohella Planck-projektissa tutkitaan myös muita radiotaivaan kohteita, kuten aktiivisia galakseja ja Linnunrataa.</p> <p>Aktiivisten galaksien säteily voi vaihdella merkittävästi jopa muutaman vuorokauden sisällä. Koska Planckin tulokset saadaan vasta lennon jälkeen, tarvitaan aktiivisten galaksien tutkimuksessa ohjelmisto, joka erottelee ja analysoi pistemäiset säteilylähteet Planckin datasta lähes reaaliaikaisesti, jolloin tulokset saadaan muutaman vuorokauden kuluttua itse havainnosta. Ohjelmiston havaintojen perusteella hälytetään muita havaintolaitoksia tarkempia jatkohavaintoja varten. Tämän diplomityön aiheena on kyseisen ohjelmiston lopullinen toteutus.</p> <p>Ohjelmisto käsittelee Planckin matalataajuusinstrumentin dataa havaintotaajuuksilla 30, 44 ja 70 GHz. Data käsitellään tunnin mittaisissa jaksoissa, joista kukin muodostaa taivaalla renkaan. Signaalin kohina vaimennetaan ja siitä erotellaan pistelähteet suodattamalla se meksikolaishattuaallokkeen määrittämällä suotimella. Löydettyjä pistelähteitä verrataan tietokannassa oleviin aikaisempiin havaintoihin ja kiinnostavista tehdään hälytys. Lopullisen päätöksen jatkohavainnoista tekee ohjelmiston operaattori. Ohjelmisto tulee toimimaan Planckin matalataajuusinstrumentin datankäsittelykeskuksessa Triestessä Italiassa koko Planck-satelliitin toiminta-ajan.</p>			
Avainsanat			
radioastronomia, signaalianalyysi, aktiiviset galaksit			

Alkusanat

Tämä diplomityö on tehty Metsähovin radiotutkimusasemalla osana Planck-satelliittiin liittyvää esityötä. Tätä kirjoitettaessa ohjelmisto on jo integroitu matalataajuusinstrumentin datankäsittelykeskukseen Triestessä.

Kiitokset kaikille Metsähovilaisille, erityisesti avusta ja neuvoista astronomiaan liittyvissä kysymyksissä.

Also thanks to Diego Herranz for help and advice on filter implementation.

30. lokakuuta 2006



Juha Aatrokoski

Sisältö

1	Johdanto	1
2	Planck	3
2.1	Kosminen mikroaaltotausta	3
2.2	Aktiiviset galaksit	4
2.3	Planck ja Metsähovi	4
2.4	Kiertorata	5
2.5	Kuvausstrategia	6
2.6	Vastaanottimet	7
2.7	Datankäsittely	7
2.8	Datan muoto	9
3	Ohjelmiston tavoitteet ja vaatimukset	11
3.1	Aikavaatimukset	11
3.2	Tarkkuus	12
3.3	Hälytyskriteerit	12
3.3.1	Ennestään tuntematon lähde	12
3.3.2	Aiempaa kirkkaampi lähde	13
3.3.3	Uusi purkaus	13
3.3.4	Lähteet, joilla on käänteinen spektri	13
4	Tieteellinen tausta	15
4.1	Pistelähteiden erottelu muusta signaalista	15
4.1.1	Pistelähteet	16
4.1.2	Suodatus	16
4.1.3	Kohinataso	19
4.1.4	Kynnysarvoistus	19
4.2	Astronomisista tietokannoista	19
5	Ohjelmiston rakenne ja toteutus	21
5.1	Perusrakenne	21
5.2	Sisällyttäminen Planck-projektiin	21
5.3	GNU Autoconf ja porttautuvuus	22
5.3.1	Ympäristövaatimukset	22

5.4	Käyttöliittymä	23
5.5	Konfigurointi	23
5.6	Ohjelman tilan tallentaminen	25
5.7	Datankäsittelyputki	25
5.8	Renkaiden välimuisti	26
5.9	Ohjelman tuloste	27
5.10	Tiedostojen ja datan formaatit	28
5.10.1	Lähdetietokanta	28
5.10.2	FITS-rengastietokanta	28
5.10.3	Normalisoitu rengastietokanta	29
5.10.4	Rengasvälimuisti	29
6	Datankäsittelyputki	30
6.1	Renkaiden normalisointi	30
6.2	Signaalin kohinan vaimentaminen	32
6.3	Signaaliarvojen muunto vuontiheydeksi	34
6.4	Suodatus	35
6.4.1	Suotimen vahvistuksen kompensointi	37
6.4.2	Epäjatkuvuudet suodatettavassa signaalissa	39
6.5	Signaalin kohinatason määrittäminen ja kynnysarvoistus	42
6.5.1	Testidata	43
6.5.2	Testin tulokset	45
6.5.3	Johtopäätökset	48
6.6	Huippujen etsiminen	48
6.7	Huippujen paikallistaminen	48
6.7.1	Merkityksellinen rengasväli	49
6.8	Lähteiden tunnistaminen	49
6.8.1	Osuma useampaan lähteeseen	50
6.9	Planeettojen käsittely	50
7	Testaus	52
7.1	Planck-simulaattori	52
8	Yhteenveto	54
8.1	Tavoitteiden täyttyminen	54
8.2	Tulevaisuus	55
	Kirjallisuutta	56

Lyhenteet

AGN Active Galactic Nuclei

ASCII American Standard Code for Information Interchange

COBE Cosmic Background Explorer

DPC Data Processing Centre

ERCSC Early Release Compact Source Catalog

ESA European Space Agency

FFTW Fastest Fourier Transform in the West

FITS Flexible Image Transport System

FITSIO FITS Input/Output

FWHM Full Width at Half Maximum

GCC GNU Compiler Collection

GNU Gnu's Not Unix

GPS Gigahertz Peaked Spectrum

HEALPix Hierarchical Equal Area iso-Latitude Pixelisation

HEMT High Electron Mobility Transistor

HFI High Frequency Instrument

L2 Lagrange 2 -piste

LFI Low Frequency Instrument

MHW Mexican Hat Wavelet

POSIX Portable Operating System Interface for Unix

QDS Quick Detection System

RMS Root Mean Square

SDSS Sloan Digital Sky Survey

SKICAT Sky Image Cataloguing and Analysis Tool

WMAP Wilkinson Microwave Anisotropy Probe

Luku 1

Johdanto

Euroopan avaruusjärjestö ESA:n Planck-satelliitin pääasiallinen tehtävä on kartoittaa kosmista mikroaaltotaustasäteilyä. Tätä varten se kuvaa koko taivaan kahteen kertaan usealla havaintotaajuudella. Planckin havaintotaajuuksilla on taustasäteilyn lisäksi myös muita säteilylähteitä, joiden vaikutus pitää pystyä poistamaan taustasäteilyä tutkittaessa. Nämä säteilylähteet ovat myös itsessään mielenkiintoisia tutkimuskohteita.

Teknillisen korkeakoulun Metsähovin radiotutkimusasemalla¹ tutkitaan aktiivisia galakseja. Ne säteilevät suhteellisen voimakkaasti radiotaajuuksilla ja ovat Planckin erottelukyvylle pistemäisiä. Metsähovi osallistuu Planck-projektiin osana sen ekstragalaktisten pistelähteiden työryhmää.

Yksi osa Metsähovin panosta on Planckin datavirrasta pistelähteitä nopeasti erottelevan ohjelmiston toteutus. Sen tarkoituksena on huomata Planckin havaitsemat pistelähteet nopeasti, jo muutaman päivän kuluttua havainnon tekemisestä. Planck-projektin varsinaiset tulokset käyttävät hyväkseen koko taivaan dataa, joten ne saadaan vasta lennon jälkeen. Ohjelmiston nimi on *Quick Detection System*, QDS.

Ensimmäisen version QDS:stä toteutti Mikko Parviainen diplomityönään [Pa2004]. Se oli kokoelma erillisiä ohjelmia ja skriptejä, ja sen toiminta ja ominaisuudet olivat hyvin rajoittuneita. Parviainen kirjoitti diplomityönsä valmistuttua QDS:stä yhdeksi ohjelmaksi integroidun version C++:lla, mutta tämän version ominaisuudet olivat suurin piirtein samat kuin alkuperäisenkin.

Tämän diplomityön aiheena on QDS-ohjelman jatkokehitys, joka käsittää tarpeellisten ominaisuuksien toteutuksen, dokumentoinnin ja ohjelman muuttamisen yleiskäyttöisemmäksi. Käytännössä työ käsitti lähes täydellisen ohjelmakoodin uudelleenkirjoituksen. Diplomityössä keskitytään lähinnä ohjelmiston kuvaamiseen eikä kosmisia säteilylähteitä kuvata kuten Parviainen teki omassa diplomityössään.

Luvussa 2 esitellään Planck-satelliitti ja selvitetään sen toimintaa ja siltä saadun datan muotoa, luvussa 3 kerrotaan tarkemmin mitkä ovat ohjelmiston

¹<http://kurp.tkk.fi/>

tavoitteet ja mitä vaatimuksia sen toiminnalle on asetettu, luvussa 4 kuvataan ratkaistavaan ongelmaan liittyvää tieteellistä tutkimusta ja taustaa, luvussa 5 esitetään ohjelmiston rakenne ja toteutus, luvussa 6 kuvataan ohjelmiston datankäsittelyn tärkeimmät vaiheet, luvussa 7 kuvataan ohjelmiston testausta ja Planck-simulaattoria ja luvussa 8 esitetään yhteenveto.

Luku 2

Planck

Planck on Euroopan avaruusjärjestö ESan satelliitti, jonka päätehtävä on kuvata koko taivas taajuuksilla 30–857 GHz kosmisen taustasäteilyn tutkimiseksi. Kirjoitushetken aikataulun mukaan se laukaistaan helmikuussa 2008 yhdessä Herschel-satelliitin kanssa, joka kuvaa taivasta infrapuna-alueella. Molempien kiertorata on Maa–Aurinko-parin Lagrange 2 -pisteen ympärillä, noin 1,5 miljoonan kilometrin päässä maasta.

Tässä luvussa kuvataan lyhyesti Planckiin liittyvää tähtitiedettä sekä hieman tarkemmin niitä seikkoja, jotka ovat olennaisia QDS:n kannalta. Tarkempia tietoja Planckiin liittyvästä tutkimuksesta saa Planckin tieteellisen ohjelman kuvauksesta [Bluebook].

2.1 Kosminen mikroaaltotausta

Kosminen mikroaaltotaustasäteily on havaittavissa jokaisessa taivaan kohdassa ja on peräisin ajalta noin 380 000 vuotta alkuräjähdyksen jälkeen, jolloin säteily pystyi ensi kertaa etenemään vapaasti. Sen spektri on noin 2,73 Kelvinin mustan kappaleen säteilyspektri, jonka intensiteetti on suurimmillaan noin 160 GHz:n kohdalla. Se löydettiin vuonna 1965 ja se on ollut siitä lähtien tärkein tiedonlähde maailmankaikkeuden muotoa ja sisältöä koskevissa kysymyksissä. Suurin osa tästä tiedosta on johdettavissa säteilyssä esiintyvistä pienistä anisotropioista eli poikkeamistan sen koko taivaan keskimääräisestä intensiteetistä. Planckin päätehtävä on mitata juuri näitä anisotropioita mikroaaltotaustassa.

Aiemmin koko taivaan kosmista mikroaaltotaustaa avaruudesta käsin ovat mitanneet satelliitit COBE (*Cosmic Background Explorer*, 1992) ja WMAP (*Wilkinson Microwave Anisotropy Probe*, 2003). Planckin on tarkoitus kartoittaa mikroaaltotaustan vaihtelut aiempaa suuremmalla herkkyydellä ja erottelukyvyllä. Etenkin sen polarisaatiomittausominaisuudet ovat merkittävästi aiempia satelliitteja paremmat, mikä moninkertaistaa mittauksista saatavan tiedon määrän.

2.2 Aktiiviset galaksit

Vaikka Planckin päätehtävä onkin kosmologinen, on sen data arvokasta monelle muullekin astronomian ja astrofysiikan alalle, sillä kosminen taustasäteily ei luonnollisestikaan ole ainoa asia mitä Planckin kuvaamissa taivaskartoissa näkyy. Kosmologian kannalta on olennaista selvittää, kuinka nämä muut säteilylähteet vaikuttavat kosmisen mikroaaltotaustasäteilyn mittauksiin ja kuinka niiden vaikutus voitaisiin tehokkaimmin poistaa. Toisaalta Planck myös tuottaa ainutlaatuista, uutta tietoa näiden lähteiden — kuten aktiivisten galaksien ja muiden ekstragalaktisten pistelähteiden — tutkijoille useiden korkeiden taajuuksien täydellisten radiokarttojen muodossa.

Metsähovissa tutkitaan aktiivisia galakseja. Tavallisten galaksien säteily on peräisin tähdistä, kun taas aktiivisten galaksien aktiivinen osa on niiden ydin. Englanninkielinen nimitys niille onkin *active galactic nuclei*, AGN. Aktiivisten galaksien alaluokkia ovat mm. kvaasarit, BL Lacertae-kohteet ja GPS-kohteet (*Gigahertz Peaked Spectrum*). Aktiivisen galaksin ytimessä olevan mustan aukon lähettäviltä lähtee kaksi pyörimisakselin suuntaista plasmasuuhkua, jotka säteilevät tietyllä perustasolla. Suihkuissa liikkuu ytimestä lähteviä shokkeja, jotka aiheuttavat niissä lisää säteilyä ja jotka havaitaan kohteen säteilyvuon väliaikaisena kirkastumisena. Näitä purkauksia tutkimalla pyritään saamaan parempaa tietoa aktiivisten galaksien rakenteesta ja toiminnasta. Juuri näiden purkauksien huomaaminen eri kohteissa on QDS:n pääasiallinen tehtävä.

2.3 Planck ja Metsähovi

Metsähovi on vahvasti mukana Planck-projektissa yhdessä Tuorlan observatorion¹ kanssa. Se on yksi aktiivisimpia jäseniä Planckin ekstragalaktisten radiolähteiden työryhmässä (Working Group 6), johon QDS kuuluu yhtenä osana. Metsähovin tutkijat koordinoivat useita työryhmiä ja ovat mukana suunnittelemassa sekä satelliitin lennon aikaista tutkimusta että lopullisesta tulostamateriaalista tehtävää tiedettä.

Metsähovin havainto- ja tutkimusohjelma on mukautettu Planckia ajatellen. Mukaan on otettu satoja uusia havaintokohteita, joista monista ei ole olemassa aiempia korkeiden radiotaajuuksien havaintoja. Tällä hetkellä ei siis tiedetä, miten useimmat niistä käyttäytyvät Planckin taajuuksilla ja ovatko ne havaittavissa vai eivät. Havaittujen uusien otosten perusteella voidaan kehittää malleja siitä, millaisia kohteita — ja kuinka paljon — Planck tulee havaitsemaan ja miten ne tulisi ottaa huomioon tulosten käsittelyssä.

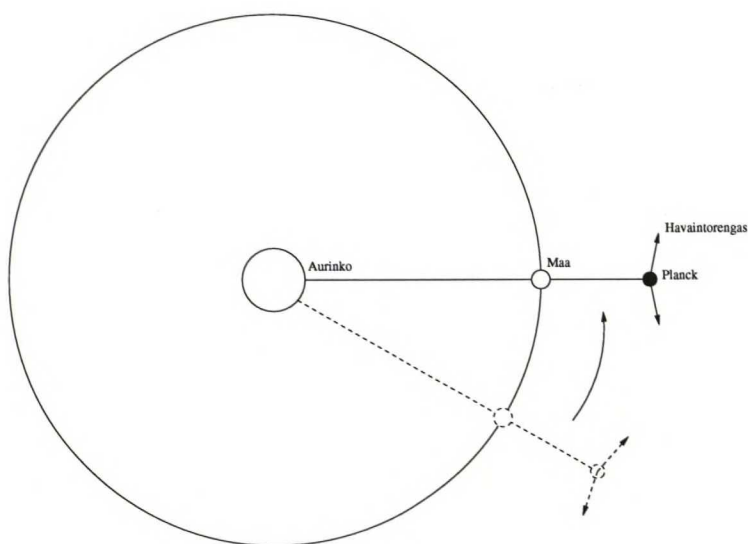
Planckin 70 GHz:n vastaanottimet on kehitetty Suomessa. Työstä vastasi Millilab, yhteistyössä Electrobit Microwaven (ent. Ylinen) sekä Metsähovin

¹<http://www.astro.utu.fi/>

kanssa. Metsähovi onkin ainoa laitos Suomessa, joka on osallistunut Planckissa sekä tieteelliseen tutkimukseen että instrumenttien rakentamiseen.

Lisäksi Metsähovin tutkijat ovat vahvasti mukana koko Suomen Planck-tiedeyhteisön toiminnassa ja koordinoinnissa. Metsähovin ja Tuorlan lisäksi Planck-tieteeseen Suomessa osallistuvat Helsingin Yliopiston Fysiikkaisten tiedeiden laitos (kosmologia) sekä Helsingin Yliopiston Observatorio (Linnunradan radiolähteet).

2.4 Kiertorata



Kuva 2.1: Planckin kiertorata (ei mittakaavassa)

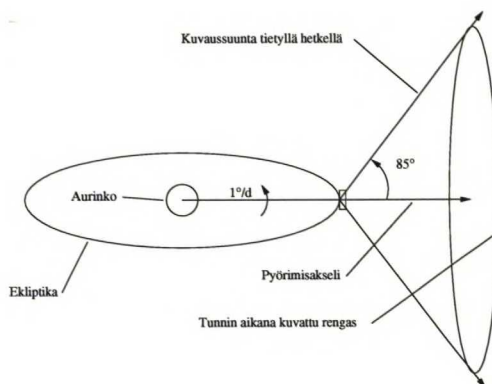
Planckin kiertorata on Maa–Aurinko-parin Lagrange 2 -pisteen (L2) ympärillä. Se sijaitsee noin 1,5 miljoonaa kilometriä Maan takana auringosta katsotuna (kuva 2.1). Vaikka Planck ei sijaitsekaan tarkalleen L2-pisteessä vaan kiertoradalla sen ympäri, voidaan taivaan kuvauksen kannalta olettaa sen sijaitsevan täsmälleen L2-pisteessä. Aurinkokunnan ulkopuolisten kohteiden kannalta voidaan niiden suuren etäisyyden vuoksi myös olettaa Planckin sijaitsevan täsmälleen joko Maan tai Auringon kohdalla, mutta koska osa aurinkokunnan kappaleista näkyy Planckin mittauksissa, täytyy myös ne ottaa huomioon (ks. kappale 6.9).

L2-kiertorata on valittu siksi, että siellä sekä Aurinko, Maa että Kuu jäävät kauas satelliitin ”taakse”, jolloin niistä peräisin olevan säteilyn haittoja on helpompi eliminoida. Lisäksi Maan kanssa viestintään käytettävä antenni osoittaa pois päin vastaanottimista, mikä myös vähentää häiriöitä.

2.5 Kuvausstrategia

Planck pyörii pääakselinsa ympäri kierroksen minuutissa. Kiertoakseli osoittaa yhteen suuntaan tunnin ajan, jonka jälkeen sitä käännetään noin 2,5 kaariminuuttia Maan kiertosuuntaan. Kiertoakseli siis kääntyy tunnin välein noin asteen päivässä eli kokonaisen kierroksen vuodessa, joten se pysyttelee samansuuntaisena Maa–Aurinko-akselin kanssa.

Vastaanottimien suuntaus on pyörimisakseliin nähden noin 85 asteen kulmassa. Satelliitin pyöriessä vastaanottimet siis kuvaavat renkaanmuotoista osaa taivaasta, *taivasympyrää*. Koska kiertoakseli pysyy samana tunnin ajan, jokainen kuvattu rengas sisältää 60 kiertoa akselin ympäri. Koska vastaanottintorvet on suunnattu hieman eri suuntiin, eivät eri vastaanotinten saman tunnin aikana kuvaamat renkaat ole samalta taivasympyrältä.



Kuva 2.2: Planckin kuvausstrategia

Akselin suunnan muutoksen ja kiertoratakorjausten aikana vastaanotinten suuntausta ei pystytä määrittelemään tarpeeksi tarkasti, eivätkä ne olisi osa mitään rengasta, joten näiltä ajanjaksoilta dataa ei voida käyttää. Miten tämä otetaan huomioon QDS:ssä on vielä auki.

Kuvausstrategiasta johtuen tulee koko taivas — teoreettisesti — kuvattua hieman yli puolen vuoden aikana. Tarkoitus on kuvata koko taivas ainakin kaksi kertaa, johon siis kuluu aikaa yli vuosi. Jotkut osat taivasta tulevat myöskin kuvattua useampaan kertaan kuin toiset. Ekliptikalla saman vastaanotinryhmän renkaat eivät juuri mene päällekkäin, kun taas napojen lähellä päällekkäisyyttä on hyvinkin paljon. Koska vastaanottimien suuntauksen kulma kiertoakseliin on alle 90 astetta, jää navoille myös täysin kuvaamattomat kalotit. Peittoa navoilla voidaan parantaa lisäämällä tunneittain tapahtuvaan akselin kääntöön myös ekliptikasta poikkeava komponentti.

QDS käsittelee Planckin dataa renkaina niiden järjestysnumeron perusteella, joka puolestaan on suoraan verrannollinen kuluneiden tuntien määrään kuvauksen aloittamisesta.

2.6 Vastaanottimet

Planckin vastaanottimet on jaettu matalataajuusinstrumenttiin (*Low Frequency Instrument*, LFI) ja korkeataajuusinstrumenttiin (*High Frequency Instrument*, HFI). LFI käsittää havaintotaajuuksia 30, 44 ja 70 GHz, HFI taajuuksia 100, 143, 217, 353, 545 ja 857 GHz. Näin monella taajuudella saadaan hyvä peitto taustasäteilyn spektrin voimakkaimmalta alueelta, mikä parantaa saatavien tulosten tarkkuutta ja helpottaa etualan lähteiden poistamista taustasäteilykartoista. LFI:ssä käytetään HEMT-radiometrivastaanottimia ja HFI:ssä bolometrivastaanottimia.

Planckin matalataajuusinstrumentin syöttötorvista kukin syöttää kahta vastaanotinta, joiden ominaisuudet eroavat lähinnä polarisaatiossa, joka on saman torven vastaanotinten kesken ristikkäin. Täten syöttötorvien lukumäärä on puolet vastaanottimien lukumäärästä. Vastaanottimien ominaisuuksia on listattu taulukossa 2.1. Syöttötorvien suuntaus näkyy kuvassa 2.3, jossa näkyy myös, että satelliitin kiertosuunnan mukaisesti samalla renkaalla on useampi torvi. Samalla renkaalla olevien torvien data voidaan yhdistää, eli taajuuksille 30, 44 ja 70 GHz saadaan siis vastaavasti 1, 2 ja 3 vastaanotinryhmää.

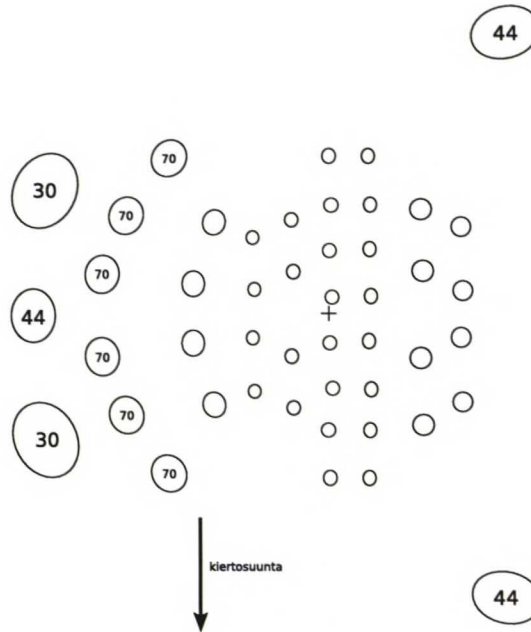
ν	N	f_s	n	s	ω	ω_s
30 GHz	4	32,5 Hz	1950	11'	33'	3,00
44 GHz	6	45,0 Hz	2700	8'	24'	3,00
70 GHz	12	76,8 Hz	4608	4,7'	14'	2,98

Taulukko 2.1: Planckin matalataajuusinstrumentin (LFI) vastaanotinten ominaisuuksia: havaintotaajuus ν , vastaanotinten lukumäärä N , näytteenottotaajuus f_s , näytteiden määrä renkaassa n , yhden näytteen leveys kaariminuuteissa s , keskimääräinen keilan puoliarvoleveys (FWHM) sekä kaariminuuteissa ω että näytteissä ω_s .

2.7 Datankäsittely

Planck toimii pääosin autonomisesti. Jokaisena vuorokautena on noin kolmen tunnin kommunikaatiojakso, jonka aikana edellisenä vuorokautena tallennettu data lähetetään maa-asemalle ja jonka aikana satelliitin toimintaa voidaan säätää. Maa-asemalta data lähetetään edelleen datankäsittelykeskuksille (*Data Processing Centre*, DPC). Vuorokaudessa Maahan lähetettävän datan määrä on luokkaa kymmeniä gigatavuja.

Datankäsittelykeskuksien sisäinen organisaatio koostuu viidestä *tasosta*. Planckin Maahan lähettämä data käsitellään neljässä vaiheessa tasoilla 1–4. Näiden lisäksi on vielä Taso S, joka käsittelee mahdollisimman tarkan simulaattorin toteuttamisen Planck-lennolle. Planck-simulaattoria on kuvattu tarkemmin kappaleessa 7.1.



Kuva 2.3: Planckin LFI-vastaanotintorvien suuntaus, ristin kohdalla nominaalinen 85° kuvaussuunta. Kuvassa näkyvät myös HFI-vastaanottimien syöttötorvet.

LFI:n ja HFI:n datan käsittely tapahtuu erillisissä datankäsittelykeskuksissa ainakin kahden ensimmäisen tason osalta; kaksi jälkimmäistä tasoa vaativat ainakin jonkinasteista yhteistyötä. Tämä on yksi syy siihen, miksi QDS toimii ainoastaan LFI:n taajuuksilla. Datan saaminen molemmista instrumenteista ei tosin olisi teknisesti kovin vaikeaa, ongelmat ja rajoitukset ovat lähinnä poliittisia ja hallinnollisia. Planckin datankäsittelyn tasot on kuvattu seuraavassa listassa.

Taso 1

Tasolla 1 erotellaan tieteellinen ja telemetriadata toisistaan. Molemmille tehdään nopeita tarkistuksia satelliitin oikean toiminnan varmistamiseksi. Data myöskin järjestetään aikajärjestykseen ja viallinen data merkitään. Taso 1 on hyvin nopea vaihe, sen kesto on tunteja tai korkeintaan päiviä.

Taso 2

Tasolla 2 tehdään datan kalibrointi sekä häiriöiden ja kohinan poisto, ja lopuksi data koostetaan taivaskartoiksi. Kalibrointia ja osaa kohinan poistosta tehdään koko ajan, mutta taivaskarttojen luontia ja $1/f$ -kohinan poistoa varten tarvitaan koko taivaan data, joka on saatavilla vasta lennon jälkeen.

QDS:n on tarkoitus käyttää niin kalibroitua dataa kuin mahdollista, mut-

ta kuitenkin niin, että data on saatavilla QDS:lle viimeistään viikon kulu-
tuttua havainnoinnista.

Taso 3

Tasolla 3 tason 2 taivaskartat erotellaan komponenttikartoiksi säteily-
lähteen perusteella. Tarkoitus on erotella omiksi kartoikseen ainakin itse
mikroaaltotausta, pistelähteet, Sunyaev–Zeldovich-efekti ja Linnunrata.
Tällä tasolla koostetaan myös pistelähdeluetteloita, joista merkittävin on
nopeasti julkaistava *Early Release Compact Source Catalog*, ERCSC. Se
julkaistaan yhdeksän kuukauden kuluttua siitä, kun koko taivas on saatu
kuvattua ensimmäisen kerran.

Taso 4

Tasolla 4 kerätyistä datasta tehdään ja arkistoidaan lopulliset julkaistavat
tulokset. Näitä ovat kalibroitu aikajärjestetty data, taivas- ja komponent-
tikartat kultakin taajuudelta sekä lopullinen pistelähdeluettelo. Tulokset
julkaistaan kaksi vuotta lennon loppumisen jälkeen. Tämä on suunniteltu
sitte, että datan jalostamiselle käytettävään muotoon on aikaa yksi vuo-
si ja tuloksien hyödyntämiselle yksinoikeudella Planck-projektin sisällä
toinen vuosi.

Tasot 3 ja 4 tarvitsevat koko taivaan datan, samoin tason 2 “loppuvaihe”,
joten ne eivät liity juurikaan QDS:n toimintaan. Pistelähdeluettelon koosta-
misessa saatetaan tosin käyttää hyväksi QDS:n tuloksia. QDS käyttää niin
hyvin kalibroitua dataa kuin on mahdollista saada sen aikarajoitusten puit-
teissa. Käytännössä tämä tarkoittaa, että QDS:n käyttämä data otetaan ta-
son 2 alkuvaiheesta, jossa kalibroitua ja häiriönpoistoa on tehty jo ainakin
jonkin verran.

2.8 Datan muoto

Simulaattorin tuottama data on FITS-tiedostojen (*Flexible Image Transport
System* [Ha2001]) muodossa. Datankäsittelykeskukselta saatavan datan muo-
toa ei ole vielä tähän mennessä tarkkaan päätetty. Se saattaa olla saatavissa
jonkinlaisesta tietokannasta, tai se saattaa olla kokoelma FITS-tiedostoja si-
mulaattorin tapaan.

Data sisältää jokaista näytettä kohden vastaanottimen antennilämpötilan
(itse näyte) sekä vastaanottimen suuntauksen ekliptisinä kulmakoordinaattei-
na. Näiden lisäksi aikaleima olisi hyödyllinen, mutta QDS:n vaatimalla aika-
tarkkuudella aikaleima voidaan laskea renkaan järjestysnumeron ja kuvauksen
aloitusajan perusteella.

Datankäsittelyn tasoilla 2–4 tehdyt taivaskartat ovat “kaksiulotteisessa”
muodossa. Koko taivaan pallopinnan pikselöimiseen eli esittämiseen joukko-

na vierekkäisiä näytteitä on useita tapoja. Järkevän pikselöinnin tärkein vaatimus on, että kaikilla pallopinnan pikseleillä on likimäärin sama muoto ja pinta-ala. Planckin taivaskartoissa käytetään HEALPix-pikselöintiä [G62005], jossa pikselit ovat samansuuruisia, niillä on hierarkkinen järjestys ja ne ovat pallopinnalla leveysasteiden suuntaisissa renkaissa.

Luku 3

Ohjelmiston tavoitteet ja vaatimukset

Planck-lennon varsinaiset tulokset tulevat vasta useita kuukausia lennon päätymisen jälkeen. Tämä johtuu pääasiassa siitä, että varsinaiset tulokset vaativat mahdollisimman suurta tarkkuutta jotta esimerkiksi komponenttierottelu sekä häiriöiden ja kohinan poisto voidaan tehdä hyvin. Tämä tarkoittaa, että käytettävissä on käytännössä oltava koko taivaan data.

Metsähovissa puolestaan ollaan kiinnostuneita vain pistelähteissä tapahtuvista nopeista säteilyvuon muutoksista. Juuri tästä syystä QDS on olemassa: haluttiin ohjelmisto, joka lähes reaaliaikaisesti löytää pistelähteet Planckin datasta ja pääättelee, onko niissä tapahtunut jotain kiinnostavaa.

3.1 Aikavaatimukset

Edellä mainittu ”reaaliaikaisuus” on määritelty QDS:n tapauksessa siten, että ohjelmiston tulisi saada Planckin data viimeistään viikon kuluttua sen havaitsemisesta. Tähän tulee lisäksi vielä QDS:n ajon vaatima aika päälle, mutta se on vain joitain minuutteja, korkeintaan tunteja. Tämä aikaraja on valittu sillä perusteella, että Planckin LFI-taajuuksilla nopeimmat aktiivisten galaksien vaihtelun aikaskaalat ovat nopeita, jopa luokkaa päiviä tai viikkoja [Lä1999], ja jatkohavainnot useilla taajuuksilla halutaan tehdä ennen kuin kohteen tila muuttuu merkittävästi.

Tämän vaatimuksen toteutuminen on kiinni Planckin datankäsittelykeskuksen toiminnasta. Ensimmäisen tason tuottama data pitäisi olla saatavilla hyvinkin nopeasti, toisen tason alkuvaiheen kalibroinnit ja häiriöiden poistot hidastavat saantia toistaiseksi tuntemattoman määrän. Mikäli datan saanti viivästyisi yli viikon aikarajan, jouduttaneen joitain käsittelyvaiheita jättämään pois. Mikäli lisäksi datan saanti ei onnistu automaattisesti vaan vaatii ihmistyötä, viivästyy se entisestään. Tämän hetkisten arvioiden perusteella viikon aikaraja näyttäisi olevan hyvin saavutettavissa.

3.2 Tarkkuus

Vuotarkkuudesta on esitetty, että virhe olisi korkeintaan 30%. Tämä tarkoittaa käytännössä, että pistelähteiden havaintokynnys on vähintään $3,33\sigma$, missä σ on suodatetun signaalin kohinataso. Näin iso virhe on hyväksyttävä, koska tässä vaiheessa halutaan lähinnä hälytys kiinnostavista tapahtumista ja tarkemmat havainnot voidaan tehdä myöhemmin muilla havaintovälineillä. Ainoa vuotarkkuuteen käytännössä vaikuttava tekijä on käytettävän signaalin (RMS-)kohinataso suodatuksen ja muun käsittelyn jälkeen, joten virheen suuruutena voidaan käyttää kohinatasoa. Kohinanvaimennusta on käsitelty kohdassa 6.2.

Paikkatarkkuuden tulisi olla sitä luokkaa, että pistelähteet voidaan erottaa toisistaan ja niitä voidaan verrata tietokannassa oleviin. Tarkkuuden on siis oltava keilan puoliarvoleveyden luokkaa, minkä ei pitäisi olla Planckin kannalta ongelma. QDS:n paikkatarkkuus voi teoriassa olla *korkeintaan* puoli näytteenleveyttä, mikä taulukon 2.1 perusteella antaa LFI-vastaanottimille paikkatarkkuudet 5,5, 4,0 ja 2,3 kaariminuuttia. Käytännössä käytetyt menetelmät saatavat aiheuttaa lisää epätarkkuutta, joten käyttökelpoinen arvio pistelähteiden paikan tarkkuudesta lienee noin yhdestä kahteen näytettä eli 5–15 kaariminuuttia. Tämän ei pitäisi aiheuttaa ongelmia, koska se on samaa suuruusluokkaa vastaanottimien keilanleveyksien kanssa.

3.3 Hälytyskriteerit

Kaikki havaitut pistelähteet eivät välttämättä ole kiinnostavia. Esimerkiksi tunnettu lähde, jonka havaittu vuo ei ole muuttunut edellisestä havaintokerrasta, ei ole kiinnostava. Kiinnostavia ovat nopeat tai suuret vuonmuutokset. Tarvitaan siis kriteerejä, joiden perusteella kohteen havaintohistoriaa tutkimalla voidaan päättää, mitkä havainnot aiheuttavat ohjelmiston operaattorille hälytyksen ja mitkä vain tallennetaan hiljaa tietokantaan ja logiin.

3.3.1 Ennestään tuntematon lähde

Mikäli havaittu pistelähde ei löydy QDS:n tietokannasta, on kyseessä uusi lähde, joka luokitellaan mielenkiintoiseksi pelkästään sen uutuuden perusteella. Tällaisia ovat myös ne lähteet, joiden ei oletettu olevan Planckin havaittavissa ja joita ei siksi otettu mukaan QDS:n tietokantaan. Luultavasti suurin osa QDS:n havaitsemista lähteistä tulee olemaan uusia, ja suuri osa näistä kohinasta, Linnunradasta tai muusta taustasta johtuvia virrehavaintoja.

3.3.2 Aiempaa kirkkaampi lähde

Lähteen vuo on tarpeeksi paljon suurempi kuin *mikään* sen historiasta löytyvä havainto kyseisellä taajuudella. Käytännössä tämä ilmaistaan epäyhtälön

$$S > cS_{max} \quad (3.1)$$

toteutumisella, missä S on nykyinen vuo, S_{max} suurin vuo samalla taajuudella ja c kynnyskerroin. Tyypillisesti $c > 1$.

Tällä kriteerillä yritetään havaita purkaus sellaisissa lähteissä, jotka eivät ole ennen purkautuneet. Tällöin purkaus tarkoittaa juurikin vuon nousua lähteen normaalin maksimitason yläpuolelle.

3.3.3 Uusi purkaus

Vuo on tarpeeksi paljon suurempi kuin *edellinen* havainto kyseisellä taajuudella, edellisen havainnon ollessa tarpeeksi tuore. Tämä ilmaistaan epäyhtälön

$$S > cS_{prev} \wedge t - t_{prev} < T \quad (3.2)$$

toteutumisella, missä S_{prev} on edellinen vuo, t_{prev} edellisen vuon ajankohta ja T aikaraja, tyypillisesti luokkaa pari kuukautta.

Tällä kriteerillä havaitaan lähteitä, joiden vuo on suhteellisen nopeassa nousussa. Tämä saattaa johtua uudesta alkavasta purkauksesta.

3.3.4 Lähteet, joilla on käänteinen spektri

Säteilylähteen *spektri-indeksi* kuvaa lähteestä mitatun vuontiheyden riippuvuutta havaintotaajuudesta. Se voidaan määrittää kahdella eri taajuudella tehdyistä havainnoista, eikä se välttämättä ole sama kaikille taajuuspareille. Spektri-indeksillä on kaksi yleistä määritelmää: taajuuteen verrannollinen ($S \propto \nu^\alpha$) ja aallonpituuteen verrannollinen ($S \propto \lambda^\alpha \iff S \propto \nu^{-\alpha}$). QDS käyttää näistä ensimmäistä, joka on yleisempi radioastronomiassa. Lähteen spektri-indeksi α taajuuksien ν_1 ja ν_2 välillä saadaan kaavalla

$$\alpha = \log \frac{S_1}{S_2} / \log \frac{\nu_1}{\nu_2}. \quad (3.3)$$

Kiinnostavia ovat sellaiset lähteet, joiden spektri-indeksi on riittävän korkea, tyypillisesti $\alpha > 1$. Tämä tarkoittaa sitä, että vuontiheys kasvaa suuremmilla taajuuksilla.

Spektri-indeksin järkevä laskeminen vaatii, että vuohavainnot ovat tarpeeksi samanaikaisia: mikäli kohde on ehtinyt havaintojen välissä muuttua, ei spektri-indeksi oikeastaan enää kerro mitään hyödyllistä. Tämän takia uutta havaintoa käsitellessä spektri-indeksin laskemiseen käytettäviä aikaisempia havaintoja pitää rajoittaa ajallisesti tarpeeksi lähelle.

Edellä kuvatuista ehdoista saadaan kriteeriksi epäyhtälö

$$\alpha_{max} > A \wedge |t_1 - t_2| < T, \quad (3.4)$$

missä T on havaintojen samanaikaisuusraja, tyypillisesti luokkaa muutama viikko.

Luku 4

Tieteellinen tausta

4.1 Pistelähteiden erottelu muusta signaalista

QDS:n varsinainen tehtävä on erotella pistelähteiden signaali kaikesta muusta signaalista. Teknisesti siis pyritään maksimoimaan signaali-kohinasuhde tapauksessa, jossa pistelähteet muodostavat signaalin ja kaikki muu on kohinaa, riippumatta tämän muun tarkasta luonteesta.

Temporaalista eli ”aitoa” kohinaa vähennetään yleensä kasvattamalla havainnoinnin integraatioaikaa. Tämä tarkoittaa, että yhtä pikseliä tai mittauspistettä havainnoidaan pidemmän aikaa joko yhtäjaksoisesti tai vaihtoehtoisesti useammassa jaksossa, joiden tulokset lopuksi keskiarvoistetaan. Planckissa käytetään jälkimmäistä tapaa: jokainen taivasympyrä kuvataan kuusikymmentä kertaa ja kohinavaimennettu rengas saadaan keskiarvoistamalla nämä kuvausrenkaat. Efektiiivistä integraatioaikaa voidaan kasvattaa myös käyttämällä useampaa instrumenttia joilla on tarpeeksi samanlaiset ominaisuudet, ja keskiarvoistamalla niiden tulokset yhteen (ks. kappale 6.2).

QDS:n tapauksessa ja monissa muissakin tapauksissa ei olla kiinnostuneita tietystä taivaan pisteestä mitatusta absoluuttisesta vuontiheydestä, vaan kiinnostavaa on se, kuinka paljon pisteen vuontiheys eroaa kyseisen kohdan taustan vuontiheydestä. Toisin sanoen Planckin mittaaman signaalin keskitasolla ei ole vaikutusta QDS:n toiminnan kannalta. Tämä on merkittävä ominaisuus, sillä laitteistosta aiheutuva $1/f$ -kohina näkyy yhdessä renkaassa keskitason siirtymisenä. Muodostettaessa kaksiulotteisia taivaskarttoja renkaiden eri tasot näkyvät ”raitoina”, ja tasoerot pitää eliminoida jotta vierekkäiset ja risteävät renkaat olisivat yhdenmukaisia keskenään [Ma2002]. Tähän tarvitaan käytännössä kaikki saatavilla oleva data, joten menetelmää ei voisi soveltaa QDS:ssä.

4.1.1 Pistelähteet

Ideaalitapauksessa pistelähde olisi juuri nimensä mukainen, eli se näkyisi ainoastaan yhdessä signaalinäytteessä. Vastaanottimen keila ei kuitenkaan ole pistemäinen, ja koska peräkkäisten näytteiden etäisyys renkaassa on pienempi kuin vastaanottimen keilanleveys, näkyy pistemäinen lähde useammassa näytteessä. Keila on yksiulotteiselta muodoltaan Gaussin käyrä, joten pistelähteet näkyvät signaalissa gaussisina huippuina joilla on keilanleveyden määrittämä leveys. Tästä seuraa se, että kaikki kompaktit säteilylähteet joiden leveys on riittävän paljon pienempi kuin vastaanottimen keilanleveys ovat ekvivalentteja "aitojen" pistelähteiden kanssa.

4.1.2 Suodatus

Temporaalisen kohinan lisäksi Planckin signaalissa on tietenkin myös tausta, eli kaikki ne näytteet, jotka eivät ole pistelähteiden aiheuttamia. Pistelähteiden erottelun kannalta tämä on myöskin kohinaa, tosin yleensä vain tarpeeksi korkeataajuisia taustasignaalin komponentteja kutsutaan kohinaksi. Taustan voidaan katsoa olevan ajan suhteen vakio, joten siitä aiheutuvaa virhettä ei voida poistaa kasvattamalla integraatioaikaa vaan se täytyy tehdä suodattamalla. Luonnollisesti suodatuksella pyritään myös vähentämään jäljellä olevaa temporaalista kohinaa.

Planck-projektissa pistelähteiden ja muidenkin komponenttien erottelua kosmisesta mikroaaltotaustasta on tutkinut pääasiassa Instituto de Física de Cantabrian tutkimusryhmä Santanderissa Espanjassa¹. Diego Herranz tästä ryhmästä on antanut neuvoja QDS:n suodatukseen liittyvien osien toteutuksessa.

Gaussista suodinta on käytetty hyvin kauan poistamaan valkoista kohinaa astronomisista kuvista. Se on kuitenkin tyypiltään alipäästösuodin, joten se ei poista taustan isompia rakenteita vaan ainoastaan korkeataajuisen kohinan. Pistelähteiden erotteluun käytettävän suotimen tulee olla tyypiltään kaistanpäästösuodin.

Santanderin tutkimusryhmä on tutkinut erilaisten suotimien soveltuvuutta pistelähteiden erotteluun sekä yksi- että kaksiulotteisissa tapauksissa. Tutkimusten pohjalta parhaiten soveltuvia ovat kolme suodinta: signaalinkäsittelyssä hyvin tunnettu optimoitu suodin (*matched filter*), meksikolaishattuaallokkeen määrittämä suodin (*Mexican hat wavelet*, *MHW*, [Vi2001, Vi2003]) ja skaalamukautuva suodin (*scale-adaptive filter*, *optimal adaptive filter*, *optimal pseudo-filter*, [Sa2001, He2002a, He2002b]). Näistä ensimmäinen ja viimeinen sovitetaan käsiteltävään signaaliin ja haettavien pistelähteiden muotoon, joten ne ovat laskennallisesti raskaita. Meksikolaishattusuodin puolestaan on hyvin yksinkertainen mutta ei yhtä optimoitu juuri tietylle signaalille.

¹<http://www.ifca.unican.es/Top/microondas.htm>

Suotimien keskinäinen paremmuus riippuu paljolti käsiteltävän signaalin taustasta, mikä ilmenee eri suotimia vertailevista tutkimuksista [He2002a, He2002b, Ba2003, Ló2005, Ló2006]. QDS:n tapauksessa asiaa mutkistaa lisäksi se, että Planckin kuvausrenkaat ovat poikittain Linnunrataan nähden. Linnunrata näkyy siis jokaisessa renkaassa kahtena kohtana (ks. kuva 6.3), joilla on erilaiset taustaominaisuudet. Jos optimoiudun ja skaalamukautuvan suotimen vaatima tausta-analyysi tehdään koko renkaasta, ei tulos välttämättä ole optimaalinen.

QDS:ssä suodatukseen käytetään meksikolaishattusuodinta, koska sillä on yksinkertainen analyttinen muoto ja se on lähes yhtä hyvä kuin optimoitu tai skaalamukautuva suodin. Lisäksi se kestää niitä paremmin systemaattisia virheitä ja vaihteluja signaalin taustan ominaisuuksissa. Suunnitelmissa tosin on toteuttaa tuki myös optimoidulle suotimelle ja testata, tuoko sen käyttö merkittävää parannusta meksikolaishattusuotimeen nähden QDS:n tapauksessa.

Eri säteilylähdekomponenttien erotteluun signaalista on olemassa muitakin tapoja kuin suodatus, kuten esimerkiksi maksimientropiametodi [Ba2004]. Nämä ovat kuitenkin monimutkaisempia ja niiden vaatimukset datan suhteen ovat erilaiset.

Meksikolaishattusuodin

Meksikolaishattuaaloke on normalisoitu Gaussin funktion toinen derivaatta, minkä johdosta se on lähes optimaalinen suodin havaitsemaan gaussisella keilalla havaittuja pistelähteitä [Sa2001]. Normalisoidun toisen derivaatan ottaminen voidaan toistaa iteratiivisesti saadulle aallokkeelle, jolloin saadaan ”perhe” eriasteisia meksikolaishattuaallokkeita [Go2006]. Tuki näille korkeamman asteen meksikolaishattuaallokkeille on suhteellisen triviaalia toteuttaa QDS:ään, ja näin onkin tarkoitus tehdä. Kaksiulotteisessa tapauksessa toisen asteen meksikolaishattuaaloke on osoittautunut ensimmäisen asteen aalloketta paremmaksi [Ló2006], mutta asian laita yksiulotteisessa tapauksessa vaatii lisäselvitystä.

Meksikolaishattusuotimen impulssivaste aika-avaruudessa on muotoa

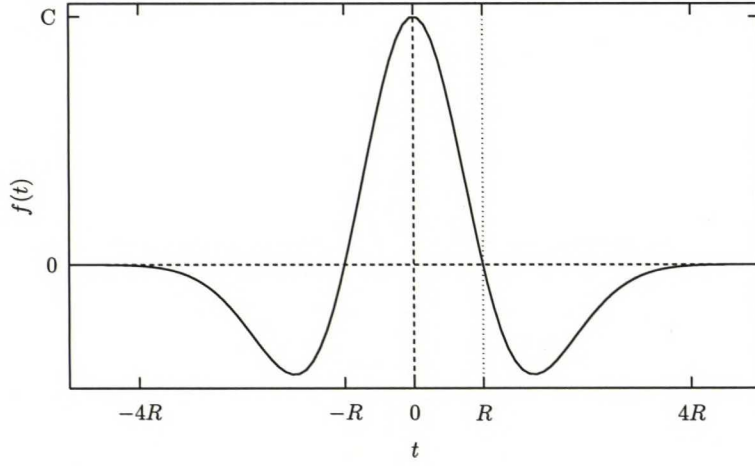
$$f(t) = C \left(1 - \left(\frac{t}{R} \right)^2 \right) e^{-\frac{1}{2} \left(\frac{t}{R} \right)^2}, \quad (4.1)$$

missä R on suotimen leveys (ks. kuva 4.1). Usein yhtälö on kirjoitettu muodossa $C = 1/\sqrt{2\pi}$.

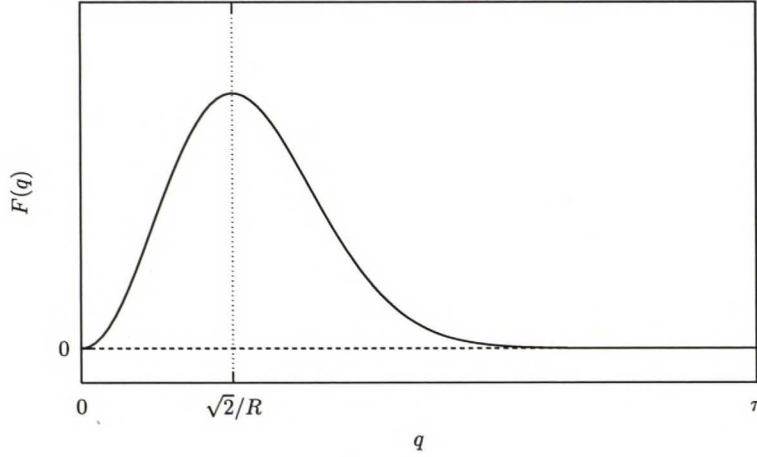
Suotimen taajuusvaste Fourier-avaruudessa on muotoa

$$F(q) = C\sqrt{2\pi}R(Rq)^2 e^{-\frac{1}{2}(Rq)^2} \quad (4.2)$$

missä $q \in [0, \pi]$ on taajuuskomponentti siten, että kohta $q = \pi$ vastaa Nyquistin taajuutta, ja C sama skaalausvakio kuin impulssivasteen yhtälössäkin. Taajuusvaste on esitetty kuvassa 4.2.



Kuva 4.1: Meksikolaishattusuotimen impulssivaste



Kuva 4.2: Meksikolaishattusuotimen taajuusvaste. Kohta $q = \pi$ vastaa Nyquistin taajuutta.

Meksikolaishatun leveysparametri R on luonnollisessa tapauksessa sama kuin gaussisen lähteen leveysparametri σ . Artikkelissa [Vi2001] kuitenkin todetaan, että optimaalinen leveys ei välttämättä ole tämä luonnollinen leveys eikä edes vakio, vaan se riippuu taustan ominaisuuksista.

Periaatteessa meksikolaishatun optimaalisen leveyden voi määrittää jokaiselle suodatettavalle renkaalle erikseen, ja tuki tälle onkin suunnitteilla toteuttaa QDS:ään. Tämä on kuitenkin huomattavasti raskaampaa kuin ennalta määrätyn leveyden käyttö, joten tarvitaan testausta selvittämään, onko optimaalisen leveyden määrittämisestä jokaiselle renkaalle käytännön hyötyä QDS:n tapauksessa.

4.1.3 Kohinataso

Kynnysarvoistamista ja havaintojen virheen suuruutta varten jokaisesta suodatetusta renkaasta pitää selvittää sen kohinataso. Vain kynnysarvon ylittävät huiput signaalissa tulkitaan havaintoina, ja kynnysarvo annetaan yleensä kohinatason moninkertana. Havaintovirheessä taas ollaan kiinnostuneita siitä, millä välillä havaitun kohteen vuontiheys on, eli haettu arvo on $S \pm \Delta S$. Molemmissa tapauksissa kohinatasolla tarkoitetaan RMS-kohinaa eli kohinan keskihajontaa (merkitään σ , mistä syystä kohinatasoa kutsutaan usein sigmaksi).

Keskihajonnan laskenta on sinänsä suoraviivaista, varsinainen ongelma on määrittää pelkän kohinan keskihajonta signaalista, jossa on kohinan lisäksi pistelähteiden aiheuttamia huippuja. Tähän käytetään usein *sigma clipping*iksi kutsuttua menetelmää, jossa lasketaan koko signaalin keskihajonta, jonka moninkertaa käytetään havaitsemaan ja poistamaan signaalista korkeimmat huiput. Sama iteraatio toistetaan karsitulle signaalille kunnes jokin lopetusehto saavutetaan. QDS käyttää hieman erilaista, kappaleessa 6.5 kuvattua menetelmää kohinataso määrittämiseen.

4.1.4 Kynnysarvoistus

Kynnysarvoistaminen on yksinkertaisin ilmaisin erottelemaan pistelähteitä signaalista: ainoastaan tietyn kynnysarvon ylittävät huiput tulkitaan havainnoiksi. Se on myös useimmissa tapauksissa, mukaanlukien QDS, täysin riittävä ilmaisin. Pistelähteiden erotteluun voidaan myös käyttää monimutkaisempia ilmaismia, jotka ottavat huipun voimakkuuden lisäksi huomioon esimerkiksi sen kaarevuuden [Ló2005]. Nämä ovat kuitenkin huomattavasti monimutkaisempia verrattuna kynnysarvoistamiseen.

4.2 Astronomisista tietokannoista

Astronomia on tietokantajärjestelmien kannalta hyvin mielenkiintoinen alue. Astronomiset tietokannat — varsinkin koko taivaan kattavat — ovat kooltaan hyvin suuria (useita, jopa kymmeniä teratavuja), niiden data on moniulotteista (kohteiden paikat ja muut parametrit) ja dataan kohdistuvat haut saattavat olla hyvinkin eksoottisia. Suuren määrän takia datan analysointi on hankalaa käsin tehtäväksi, joten astronomia on myös hyvä sovelluskohde tiedonrikastukselle (*data mining*). Astronomiset tietokannat ovatkin yksi tärkeimpiä tietokanta- ja tiedonrikastusala eteenpäin vievistä voimista. Esimerkkinä suurista astronomisista tietokannoista annettakoon Sloan Digital Sky Survey². SDSS-tietokannan rakennetta, suunnittelua ja toteutusta kuvataan artikkeleis-

²<http://www.sdss.org/>

sa [Sz2000] ja [Sz2002]. Esimerkki tiedonrikastusjärjestelmästä on astronomisen datan automaattinen luettelointi- ja analyysityökalu SKICAT [Fa1993].

Perinteisesti astronomisen datan tallennukseen ja siirtoon on käytetty tähän tarkoitukseen suunniteltua FITS-tiedostoformaattia (*Flexible Image Transport System* [Ha2001]). Alunperin FITS oli tarkoitettu lähinnä kuvien tallentamiseen, mutta eri laajennuksien myötä se tukee monenmuotoista dataa. Lisäksi tiedostoformaattiin kuuluu kohtuullisen yksityiskohtainen ASCII-muotoinen kuvaus siitä, miten data tulee tulkita. Tämän johdosta hyvin tehtyjen FITS-tiedostojen tulkinta ei teoriassa vaadi ylimääräistä dokumentaatiota. FITS on siis alan hyväksytty ja paljon käytetty standardimuoto, mutta se ei sovellu erittäin suuriin tietokantoihin niiden suorituskykyvaatimusten takia.

Raakadatan kohdalla näitä yllä mainittuja ongelmia ei juurikaan ole, sillä raakadataa käytetään yleensä vain varsinaisten, jalostettujen tietokantojen luomiseen. Koska nämä operaatiot ovat kertaluonteisia ja harvinaisia eivät ne myöskään ole rinnakkaisia, suorituskyky ei ole niiden kannalta kovin kriittinen ominaisuus ja raakadatatietokantojen muodolla ei ole niin suurta merkitystä. (Tässä on käytetty määritelmää, joka erottelee raakadatan ja havaintodatan toisistaan. Tietyissä sovelluksissa ensimmäistä ei varsinaisesti ole edes olemassa.)

Planckin tapauksessa kohinavaimennetun rengastietokannan koko on vain kuudeskymmenesosa raakadatan koosta. Pelkkää havaintodataa sisältävä rengastietokanta ei ole mikään suuri haaste tietokantasuunnittelun kannalta. Kaksiulotteisia taivaskarttoja sisältävä tietokanta on vielä pienempi, koska siihen on yhdistetty data kahdesta koko taivaan kuvauksesta. Sekään ei ole erityisen haastava, sillä sekin sisältää pelkkää havaintodataa, nyt tosin kuvina eikä renkaina, mutta ei tarkempia parametreja esimerkiksi pistelähteistä.

QDS käyttää joko raakadataa tai kohinavaimennettua rengasdataa, joten rengastietokanta ei muodosta erityistä ongelmaa sen kannalta. QDS:n tietokannat ovat sen omia tietokantoja FITS-tiedostomuodossa (ks. kappale 5.10). QDS:ään toteutetaan myös tuki suoralle yhteydelle DPC:n rengastietokantaan, mikäli se tulee olemaan jotain muuta kuin kokoelma FITS-tiedostoja simulaattorin tapaan. Tämä voidaan kuitenkin tehdä vasta sen jälkeen, kun DPC:n tietokantamäärittely on valmis.

Luku 5

Ohjelmiston rakenne ja toteutus

5.1 Perusrakenne

Ohjelmisto koostuu yhdestä C++:lla toteutetusta ohjelmasta, johon on integroitu kaikki tarvittava toiminnallisuus. Se on unix-komentoriviohjelma, jota voi valinnaisesti käyttää myös interaktiivisessa toimintatilassa. Ohjelman toimintaa ohjataan asetustiedostolla ja komentorivioptioilla.

Toiminnallisuuden ydin on ohjelman datankäsittelyputki (kuva 5.2). Sen sisältämät operaatiot tehdään jokaiselle määritetylle datarenkaalle vastaanotinryhmittäin, ja lopuksi tulokset kootaan yhteen.

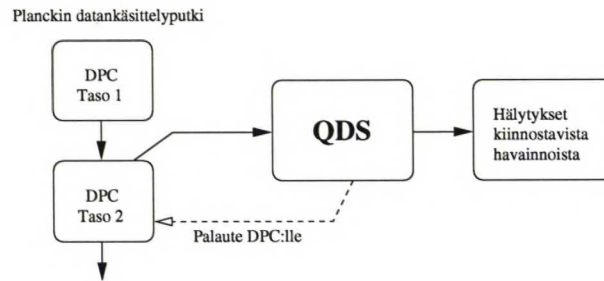
Toimiakseen ohjelma tarvitsee rengastietokannan, joka sisältää varsinaisen signaalidatan, sekä lähdetietokannan, joka sisältää tunnetut pistelähteet ja niiden havaintohistorian.

5.2 Sisällyttäminen Planck-projektiin

Pääasiallisesti QDS on tarkoitettu ja suunniteltu Metsähovin tutkimusryhmän käyttöön, koska se on kiinnostunut extragalaktisten radiolähteiden muuttuvuudesta. QDS antaa hälytyksen kiinnostavista Planckin havaitsemista pistelähteistä, minkä perusteella Metsähovi ja sen kanssa yhteistyötä tekevät muut observatoriot voivat tehdä lähteistä hyvin nopeasti tarkempia havaintoja useilla taajuuksilla. Saadut tulokset syötetään takaisin Planckin datankäsittelyputkeen, mikäli tämä katsotaan tarpeelliseksi.

QDS:n käyttämä data tulee datankäsittelyputken tason 2 alusta (kuva 5.1). Tarkemmin sanottuna datalle on tehty kaikki tason 1 peruskäsittely ja niin paljon kalibrointia ja virheenpoistoa tasolla 2 kuin on mahdollista annetulla viikon aikarajoituksella.

Periaatteessa QDS:ää voisi alustasiirrettävyytensä ansiosta ajaa missä tahansa, esimerkiksi Metsähovissa. Sille syötettävän datan määrä on kuitenkin Planckin kohdalla kohtuullisen suuri, useampi gigatavu vuorokaudessa, joten dataa ei ole välttämättä järkevää ruveta siirtämään verkon yli kovin kauas,



Kuva 5.1: QDS:n paikka Planckin datankäsittelyssä

varsinkaan useampaan kertaan. Tämän vuoksi QDS:ää tullaankin Planckin lennon aikana ajamaan LFI-datankäsittelykeskuksessa Triestessä.

5.3 GNU Autoconf ja porttautuvuus

Yksi QDS:n suunnittelutavoitteista oli tehdä siitä mahdollisimman yleiskäyttöinen ja joustava. Tähän kuuluu myös siirrettävyys eri tietokoneympäristöjen välillä eli porttautuvuus tai alustasiirrettävyys, joka yleisesti ottaen on hyvä ja toivottava ominaisuus mille tahansa ohjelmalle. Alustasiirrettävyyden parantamiseksi QDS käyttää GNU Autoconfia¹, joka huolehtii ohjelman kääntämisestä ja asennuksesta ja osaa automaattisesti huolehtia tietyistä käyttöjärjestelmäympäristöjen eroista.

Autoconfin oppimiskynnys on hieman korkea, mutta se on loistava apu alustasiirrettävien ohjelmien toteuttamisessa. Lisäksi se tekee myös valmiin ohjelman kääntämisestä ja asentamisesta helpompaa. Se ei kuitenkaan ole mikään kaiken kattava ratkaisu alustasiirrettävyyden ongelmaan. Ensinnäkin itse lähdekoodi pitää osata kirjoittaa alustasiirrettävyyttä ajatellen, mikä vaatii kohdittuullisesti tietämystä eri ympäristöjen eroista ja yleisimmistä siirrettävyysongelmistä. Toiseksi on olemassa sellaisia siirrettävyysongelmiä, joille Autoconfissa ei ole tukea tai tuki ei ole kovin hyvä tai kattava. Esimerkkinä mainittakoon POSIX-säikeisiin liittyvät asiat.

5.3.1 Ympäristövaatimukset

Seuraavassa on lueteltu QDS-ohjelmiston “viralliset” vaatimukset käännös- ja ajoympäristöltään. Nämä eivät välttämättä ole täysin tyhjentävät, mutta ne lienevät riittävät useimmissa tapauksissa.

- POSIX-yhteensopiva unix-tyyppinen käyttöjärjestelmä. QDS on kehitetty Debian/Linux-ympäristössä, mutta sen tulisi olla hyvin siirret-

¹<http://www.gnu.org/software/autoconf/>

tävissä muihin unix-tyyppisiin ympäristöihin. Siirrettävyys Windows-ympäristöön *saattaa* olla mahdollista, mahdollisesti käyttäen Cygwinia².

- Standardi C++ kääntäjä ja linkkeri, mieluummin GCC:n sisältämä g++.
- Make-ohjelma, mieluummin GNU Make.
- FFTW-kirjasto³ ja sen otsikkotiedostot, versio 3.0 tai yhteensopiva. Tätä käytetään Fourier-muunnoksiin signaalin suodatuksen yhteydessä.
- FITSIO C-kirjasto⁴ ja sen otsikkotiedostot (cfitsio), versio 2.510 tai yhteensopiva. QDS:n signaalitietokannat käyttävät tallennusmuotonaan FITS-tiedostoja.
- Curses-kirjasto ja otsikkotiedostot interaktiivista käyttöliittymää varten (valinnainen).
- X-ikkunointiympäristön asiakaskirjastot ja -otsikkotiedostot graafien piirtämistä varten (valinnainen).

5.4 Käyttöliittymä

Ohjelman pääasiallinen käyttötapa on normaalina unix-komentoriviohjelmana, jolloin ohjelman tulokset kirjoitetaan *stdout*-virtaan ja virheviestit sekä mahdolliset debug-viestit *stderr*-virtaan. Kaikki tuloste kirjoitetaan myös lokitiedostoon, mikäli sellainen on määritelty. Tulokset voidaan myös valita lähetettäväksi sähköpostilla ohjelman operaattorille.

Normaalin tavan lisäksi QDS voidaan käynnistää myös tekstipohjaiseen (curses-kirjasto) interaktiiviseen käyttöliittymään. Siinä muuttujille voi määrittää arvot tekstikentissä, ja ohjelman ajon jälkeen tulokset näkyvät valikkona, josta niitä voi tarkastella lähemmin. Lisäksi pistelähteiden havaintohistoriaa voi tarkastella graafina (toteutettu Xlib-kirjastolla), mikä helpottaa tulosten merkityksen arvioimista.

5.5 Konfigurointi

QDS:n toimintaa ohjataan useilla asetusmuuttujilla, joiden arvot määritellään ohjelman asetustiedostossa. Asetustiedosto voidaan määrittää komentorivillä,

²<http://cygwin.com/>

³*Fastest Fourier Transform in the West*, <http://www.fftw.org/>

⁴<http://heasarc.gsfc.nasa.gov/docs/software/fitsio/fitsio.html>

muutoin sitä etsitään nykyisestä hakemistosta ja sen jälkeen käyttäjän kotihakemistosta (tiedostot `./qds.conf` ja `~/qds.conf`). Asetusmuuttujan arvon voi myös antaa komentorivioptiona, jolloin se syrjäyttää asetustiedostossa olevan.

Ohjelma etsii kulloinkin tarvitsemalleen muuttujalle arvon käyttäen seuraavanlaista funktiota:

```
get(muuttujanimi, ryhmä, taajuus)
```

Tämä mahdollistaa sekä ryhmä- että taajuuskohtaiset muuttujat oletusmuuttujan lisäksi. Kun tietylle muuttujalle etsitään arvoa, käydään todellisuudessa läpi järjestyksessä seuraavat muuttujanimet ja käytetään näistä ensimmäistä, jolle on määritelty arvo:

1. **.muuttujanimi*

Ohitusarvo, joka korvaa muut määritellyt arvot. Käyttötarkoitus pääasiassa joidenkin muuttujien tilapäinen arvon muuttaminen riippumatta ryhmästä tai taajuudesta.

2. *ryhmä.muuttujanimi*

Ryhmäkohtainen arvo. Ryhmä on joko vastaanotinryhmä (pistelähteiden etsimiseen liittyvät muuttujat) tai pistelähteen tyyppi (esim. selvitetäessä hälytyskriteereitä).

3. *taajuus.muuttujanimi*

Taajuuskohtainen arvo. Tämä koskee niitä vastaanotinryhmiä, joilla on kyseinen nimellinen havaintotaajuus. Taajuus ilmoitetaan kokonaislukuina gigahertseissä.

4. *muuttujanimi*

Muuttujan oletusarvo. Käytetään, mikäli muita arvoja ei ole määritelty. Lisäksi joillain muuttujilla ryhmä- tai taajuuskohtaiset arvot eivät ole mielekkäitä (esim. lokitiedoston sijainti jne.).

Mikäli muuttujalle ei ole määritelty arvoa, sen arvoksi tulee tyhjä merkkijono, jonka numeerinen arvo on nolla.

Asetustiedosto sisältää lähinnä muuttujanimi-arvo -pareja muodossa *muuttujanimi = arvo*. Useamman ryhmä- tai taajuuskohtaisen muuttujan määrittelyn helpottamiseksi asetustiedostossa voi käyttää myös modaalista prefiksiä: rakenne

```
begin prefiksi
  nimi1 = arvo1
  nimi2 = arvo2
end
```

tuottaa saman tuloksen, kuin jos kirjoittaisi

prefiksi.nimi1 = *arvo1*

prefiksi.nimi2 = *arvo2*

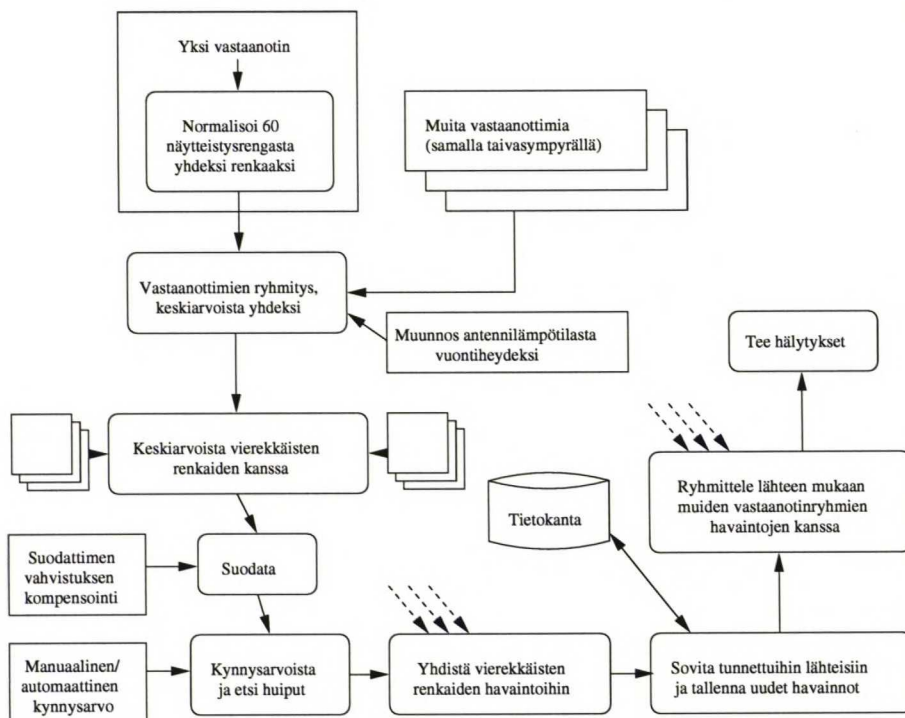
5.6 Ohjelman tilan tallentaminen

Haluttaessa QDS tallentaa viimeisen käsittelemänsä renkaan numeron tiedostoon. Antamalla sopiva komentorivioptio seuraavalla ajokerralla voidaan käsittely aloittaa automaattisesti ensimmäisestä käsittelemättömästä renkaasta ilman, että tarvitsisi muistaa mihin jäätiin edellisellä kerralla. Koska viimeisen saatavissa olevan renkaan numero selviää itse tietokannasta, tilan tallennus mahdollistaa uusien saapuneiden renkaiden käsittelyn päivittäin tai viikoittain samoilla komentoriviparametreilla, mikä tekee prosessin automatisoinnista helppoa. QDS voidaan esimerkiksi ajaa ajastettuna toimintona (cron) ja ohjata tuloste sähköpostin kautta operaattorille.

5.7 Datankäsittelyputki

QDS:n ydin on sen datankäsittelyputki. Ohjelmistolle annetaan tietty rengasväli ensimmäisen ja viimeisen käsiteltävän renkaan muodossa. Se käsittelee jokaisen renkaan numerojärjestyksessä yksi kerrallaan, minkä jälkeen se yhdistää kaikkien renkaiden tulokset. Datankäsittelyn vaiheet ovat kuvassa 5.2 ja niiden sisältö on kuvattu seuraavassa listassa. Tärkeimpien vaiheiden toteutus on kuvattu luvussa 6.

1. Rengas *normalisoidaan* keskiarvoistamalla kaikki (60) havaintorengasta yhdeksi kohinavaimennetuksi renkaaksi, joka alkaa tietyistä konsistentista kohdasta.
2. Rengas keskiarvoistetaan muiden renkaiden kanssa, jotka ovat peräisin samaan vastaanotinryhmään kuuluvilta vastaanottimilta. Tämä tapahtuu samalla tavoin kuin edellinenkin vaihe. Tässä vaiheessa myös näytteiden arvo muunnetaan antennilämpötilasta vuontiheydeksi.
3. Mikäli halutaan, rengas keskiarvoistetaan myös ympäröivien renkaiden kanssa.
4. Rengas suodatetaan. Suodatus poistaa suuret rakenteet ja korostaa pistelähteistä aiheutuvia huippuja. Suodin luodaan siten, että pistelähteistä aiheutuvilla huipuilla on suodatuksen jälkeen oikea vuontiheys.



Kuva 5.2: QDS:n datankäsittelyputki

5. Suodatetusta signaalista etsitään kaikki määritellyn kynnyksen ylittävät huiput. Kynnys määritellään joko manuaalisesti tai normaalisti kohinatason σ moninkertana.
6. Havaintoja kaikista kohderenkaista verrataan toisiinsa ja ryhmitellään yhteen samasta lähteestä peräisin olevat havainnot.
7. Verrataan saatuja havaintoryhmiä tietokantaan ja päätellään, mitä tunnettua lähdeä ne vastaavat vai ovatko ne ennestään tuntemattomia. Tietokantaan tallennetaan uusi havainto sekä luodaan tarpeen vaatiessa uusi lähde.
8. Mikäli havainto luokiteltiin mielenkiintoiseksi, siitä tehdään hälytys ohjelman operaattorille.

5.8 Renkaiden välimuisti

Suorituskyvyn parantamiseksi voidaan paikallisen levyn hakemistoa käyttää renkaiden välimuistina. Välimuistiin tallennetaan normalisoidut renkaat vastaanotinkohtaisesti painotetussa muodossa, jolloin tallennettavat kentät ovat kumulatiivinen signaali, signaalin painot sekä ekliptinen longitudi ja kolatitudi.

Itse tietokannasta lukemisen ja normalisoinnin sijasta rengas voidaan siis lukea suoraan välimuistista. Välimuistin rakenne on kuvattu kappaleessa 5.10.4.

Planckin ja muidenkin samantyyppisten signaalitietokantojen voidaan olettaa olevan muuttumattomia: kerran tietokantaan tallennettua dataa ei enää myöhemmin muuteta⁵. Tämän vuoksi välimuistiin ei ole implementoitu minkäänlaista koherenssitarkistusta, vaan kaikkien välimuistista löytyvien renkaiden oletetaan olevan ajan tasalla. QDS tarkistaa ainoastaan sen, että välimuistissa olevat renkaat ovat peräisin oikeasta tietokannasta. Renkaiden automaattista poistamista välimuistista levytilan loppuessa ei ole myöskään implementoitu.

Välimuistia ei käytetä sellaisten tietokantojen kanssa, jotka sisältävät valmiiksi normalisoitua dataa ja jotka sijaitsevat lokaalilla levyllä.

5.9 Ohjelman tuloste

Normaalina tulosteena ohjelmalla on ne havaitut pistelähteet, jotka aiheuttivat hälytyksen. Valinnaisesti tulostetaan myös ne havaitut lähteet, jotka eivät aiheuttaneet hälytystä. Ohjelman tuloste yhtä pistelähdettä kohden näyttää seuraavanlaiselta:

```
+ /! Test3 19.56° -20.58° (±27') [D30,D44-1,D70-1]
/ D30(30 GHz) 4.81 Jy ±18% @19.6,-20.6 20080102/080000 R21@31%
+! D44-1(44 GHz) 3.76 Jy ±23% @19.52,-20.55 20080102/100000 R23@31%
- D70-1(70 GHz) 3.24 Jy ±26% @19.58,-20.54 20080102/040000 R17@31%
```

Tuloste yhtä lähdettä kohden koostuu tunnisterivistä sekä sisennetyistä havaintoriveistä. Kukin rivi alkaa kokoelmalla merkkejä, jotka ilmaisevat mitä hälytyksiä kohde on aiheuttanut. Hälytysten tunnusmerkit ovat ‘*’ täysin uudelle lähteelle, ‘+’ aiempaa kirkkaammalle lähteelle, ‘/’ nopealle vuon nousulle, ‘!’ käänteiselle spektrille ja ‘-’ mikäli yksikään hälytys ei lauennut (nämä näytetään valinnaisesti).

Tunnisterivillä on ensimmäiseksi hälytysmerkkien jälkeen kohteen nimi. Mikäli kyseessä on aiemmin tuntematon kohde, nimi generoidaan automaattisesti. Seuraavaksi on kaikkien lähteestä tällä ajokerralla tehtyjen havaintojen sijainnin keskiarvo ekliptisinä koordinaatteina asteissa, sekä minkä säteen sisällä tehdyt havainnot ovat tästä keskiarvosta kaariminuuteissa (sisältää myös yhden tai kahden näytteenleveyden “luontaisen” epätarkkuuden). Koordinaatit voidaan haluttaessa näyttää myös galaktisina tai ekvatoriaalisina. Lopuksi on vielä tieto siitä, mitkä vastaanotinryhmät havaitsivat kohteen.

Havaintorivejä on yksi kutakin sellaista vastaanotinryhmää kohden, jonka käsittelyn yhteydessä kohde aiheutti hälytyksen. Havaintorivi alkaa tämän

⁵Paitsi poikkeustapauksissa, jolloin muutos luultavasti koskee koko tietokantaa ja välimuistin voi tyhjentää manuaalisesti.

vastaanotinryhmän nimellä sekä sen nimellistajuudella. Seuraavaksi tulee havaittu vuontiheys ja sen virhe (1σ kohinataso), kyseisen havainnon koordinaatit asteina, havainnon aikaleima ja lopuksi tieto missä renkaassa ja missä kohtaa kyseinen havainto sijaitsee.

5.10 Tiedostojen ja datan formaatit

5.10.1 Lähdetietokanta

Lähdetietokanta on tekstitiedosto, johon on tallennettu kaikki tunnetut lähteet, ja sinne tallennetaan myös ne QDS:n löytämät lähteet jotka eivät sieltä jo löydy. Jokaisen lähteen kohdalla on tallennettu myös sen havaintohistoria, joka koostuu havainnoista eri taajuuksilla ja johon Planckin tekemät havainnot lisätään. Lisäksi tietokanta sisältää uniikkiarvon, jonka perusteella uusille lähteille pystytään automaattisesti luomaan uniikki nimi.

Lähteestä on tallennettu sen nimi, sijainti, tyyppi sekä mahdollisesti lähdekohtaiset hälytyskriteerit (ks. kappale 3.3). Jokaisesta havainnosta on tallennettu ajankohta, havaintotaajuus, vuo ja sen virhe, havainnon tekijä sekä muutamia havainnon ominaisuuksia binäärilippuina.

Lähdetietokannan sisältö näyttää seuraavalta:

```
UNIQUE=0
Source1:0.3254289:1.28627:qso
20050101:30:5.545:0.278:Planck

Source2:3.4177:2.419036:qso
20050101:30:7.145:0.335:Planck
20050101:44:6.248:0.531:Planck:U

:
```

5.10.2 FITS-rengastietokanta

Tämä on Planckin simulaattoripaketin tulostusmuoto, joten tietokanta on tehty sen ehdoilla. Tietokanta sijaitsee jossain hakemistossa, jonka sisällä ovat vastaanotinkohtaiset FITS-datatiedostot. Yhden vastaanottimen data on jaettu kahteen tiedostoon, jotka ovat molemmat liukulukuarvoisia FITS-binääritaulukoita. Toisessa tiedostossa on vain yksi sarake, signaali, ja toisessa tiedostossa kaksi saraketta ekliptisille longitudille ja kolatitudille. Kummankin tiedoston otsakkeeseen on tallennettu sen sisältämä rengasväli sekä renkaassa olevien näytteiden määrä⁶.

⁶Lopullinen määrä normalisoinnin jälkeen; raakanäytteiden määrä voidaan päätellä taulukon rivimäärästä.

Signaalitiedoston nimi (esimerkkinä vastaanotin LFI-28b) on muotoa `signal_LFI-28b_tod_0000.fits` ja koordinaattitiedoston nimi on muotoa `signal_LFI-28b_detpt2_0000.fits`. Nämä nimet tulevat Planck-simulaattorin käyttämistä tiedostoista

5.10.3 Normalisoitu rengastietokanta

Tietokanta voi olla myös valmiiksi normalisoitu, jolloin vältetään aikaa vievä normalisointivaihe renkaiden käsittelyssä. Normalisoitu tietokanta muodostuu jossain hakemistossa olevista datatiedostoista, joiden muoto on muuten sama kuin välimuistin datatiedostojen (ks. alla), paitsi että kaikki renkaat on tallennettu järjestyksessä peräkkäin yhteen vastaanotinkohtaiseen tiedostoon eikä välistä puuttuvia renkaita sallita. Tiedoston nimi on yksinkertaisesti muotoa `LFI-28b.fits`.

Normalisoitua tietokantaa voi käyttää myös silloin, kun käsiteltävä data ei ole peräisin taivaanpallolle piirtyvästä renkaasta kuten Planckin tapauksessa. QDS:n normalisointioperaatio vaatii, että data on tällaisissa renkaissa ja että renkaan keskipiste (renkaan pisteiden muodostamassa tasossa) ei ole taivaanpallon keskipisteessä (ks. kappale 6.1). Koska normalisointivaihe jätetään tällaista tietokantaa käytettäessä pois, voi data periaatteessa koostua mielivaltaisista⁷ yksiulotteisista signaaleista.

5.10.4 Rengasvälimuisti

Välimuistihakemistossa on tekstitiedosto `QDS-CACHE-ID`, jonka sisältö on välimuistissa olevien renkaiden lähdetietokannan nimi (muodossa *tyyppi:sijainti*). Jokaisen vastaanottimen renkaat ovat vastaanottimen nimen mukaisessa alihakemistossa, ja renkaan numero on itse tiedoston nimessä. Välimuistitiedoston nimi suhteessa välimuistihakemistoon on siis muotoa `LFI-28b/ring_1.fits`.

Itse välimuistitiedostot ovat FITS-binääritaulukoita, joissa on neljä liukulukusaraketta: kumulatiivinen signaali antennilämpötilana, signaalin paino sekä ekliptinen longitudi ja kolatitudi radiaaneina. Lisäksi tiedostojen otsakkeeseen on tallennettu vastaanottimen nimi, renkaan numero ja renkaassa olevien näytteiden määrä.

Välimuistihakemistoa voi myös käyttää itse varsinaisena rengastietokantana. Tällöin normaali välimuistin toiminta on luonnollisesti kytketty pois, koska siitä ei olisi kuin haittaa ohjelman toiminnalle.

⁷QDS tosin olettaa, että ensimmäinen ja viimeinen näyte sijaitsevat rengasmaisesti vierekkäin, mutta tämä on harvoin ongelma.

Luku 6

Datankäsittelyputki

Tässä luvussa kuvataan tarkemmin tärkeimmät QDS:n datankäsittelyputken vaiheet. Datankäsittelyputken rakenne on kuvattu kappaleessa 5.7

6.1 Renkaiden normalisointi

Tässä kappaleessa käytetään seuraavaa notaatiota ja semantiikkaa vektorien kanssa:

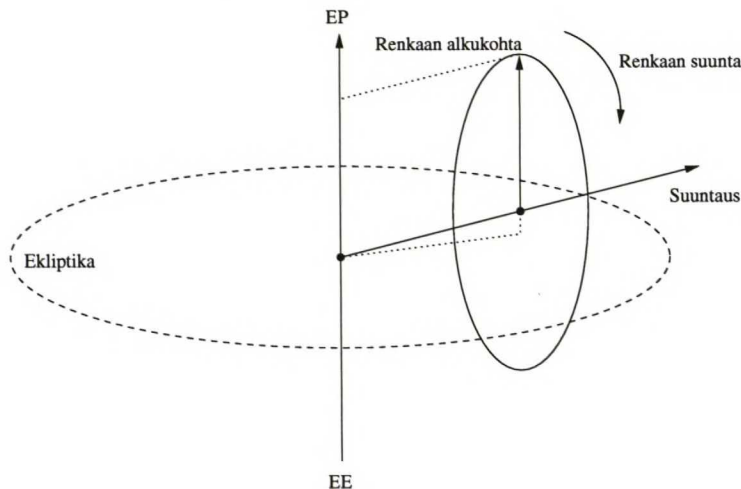
- \mathbf{v} Karteesinen yksikkövektori, jonka määrittävät komponentit v_x, v_y, v_z .
- \mathbf{v}' Karteesinen vektori joka saadaan vastaavasta (sama kirjain) yksikkövektorista skaalaamalla positiivisella kertoimella ts. $\mathbf{v}' = a\mathbf{v}$, $a > 0$.
- \vec{p} Kulmiin perustuva paikkakoordinaatti eli suuntaus, joka koostuu ekliptisestä longitudista ja kolatitudista (λ, β) . Mielletään ekvivalentiksi yksikkövektorien kanssa ts. $\vec{p} \equiv \mathbf{p}$.

Ennen kuin tietokannasta luettua renkaan dataa käytetään mihinkään, se *normalisoidaan*. Normalisoinnissa mielivaltainen määrä signaalinäytteitä ja vastaanottimien suuntauksia samalta taivasympyrältä muunnetaan yhdeksi renkaaksi, joka alkaa hyvin määritellystä pisteestä ja jossa on tietty määrä näytteitä (n). Planckin datassa jokaista rengasta vastaa 60 näytteistysrengasta, ja jokaisessa näytteistysrenkaassa on $60f_s$ näytettä (ks. taulukko 2.1), joten alkuperäisten näytteiden määrä renkaassa on $3600f_s$.

Planckin vastaanotinten erottelukyky renkaassa on noin¹ $60f_s$. Normalisoidun renkaan näytteiden määrä ei tulisi olla suurempi kuin erottelukyky, muuten osa näytteistä jää tyhjäksi alla kuvattua normalisointimenetelmää käyt-

¹Saattaa käytännössä olla suurempi, mikäli näytteistysrenkaat menevät lomittain.

täen. Koska näytteitä on hyvä olla mahdollisimman paljon, on siis Planckin tapauksessa käytetty normalisoidun renkaan näytemäärä $n = 60f_s$.



Kuva 6.1: Renkaiden normalisointi

Normalisoitu rengas alkaa siitä näytteestä, jonka sijainti on lähinnä ekliptistä pohjoisnapaa, eli siitä näytteestä, joka on positiivisen z -akselin \mathbf{z} ja vastaanottimen suuntauksen \vec{p} määrittämässä tasossa olevista näytteistä pohjoisempi. Seuraavien näytteiden paikat eli suuntaukset johdetaan kiertämällä ensimmäistä paikkaa vektorin \vec{p} ympäri positiiviseen kiertosuuntaan. Normalisointi on havainnollistettu kuvassa 6.1. Signaali- ja sijaintinäytteiden paikkaa tulosrenkaan $n:n$ kokoisessa taulukossa kutsutaan niiden *lokeroksi*.

Käytetty normalisointialgoritmi on kuvattu alla:

1. Alustetaan kaikki n paino- ja signaalidatalokeroa nollassi.
2. Selvitetään satelliitin suuntausvektori muuntamalla kaikki suuntaukset taivasympyrällä \vec{s}_i karteesisiksi \mathbf{s}_i ja laskemalla niiden keskiarvo, joka on etsitty suuntaus \mathbf{p}' (suuntaus \vec{p} on ekvivalentti normalisoidun \mathbf{p} kanssa).
3. Selvitetään "nollavektori" \mathbf{o} , joka on kohtisuorassa vektoriin \mathbf{p} , tasossa \mathbf{pz} ja osoittaa positiivisen z -akselin suuntaan. Normalisoimatonta nollavektoria \mathbf{o}' kuvaa rajoite $\mathbf{o}' \cdot \mathbf{z} > 0 \wedge \mathbf{o}' \cdot \mathbf{p} = 0 \wedge \mathbf{o}' \cdot (\mathbf{p} \times \mathbf{z}) = 0$. Koska vaatimus tälle konseptille on $\mathbf{p} \nparallel \mathbf{z}$ ja siis $p_x \neq 0 \vee p_y \neq 0$, vektori \mathbf{o}' saadaan asettamalla $o'_x = -p_z p_x \wedge o'_y = -p_z p_y \wedge o'_z = -p_z^2 + 1$ josta lopuksi saadaan $\mathbf{o} = \mathbf{o}' / |\mathbf{o}'|$.
4. Jokaista signaali- (d_i) ja sijaintinäytettä (\mathbf{s}_i) kohden tehdään seuraava:
 - (a) Projisoidaan sijaintivektori \mathbf{s}_i taivasympyrän tasoon, saadaan $\mathbf{v}' = \mathbf{s}_i - \mathbf{p}'$.

- (b) Lasketaan vektoreiden \mathbf{o} ja \mathbf{v}' välinen kulma $\psi = \arccos(\mathbf{o} \cdot \mathbf{v})$ ja si-
joitetaan se välille $[0, 2\pi)$ vertaamalla vektoreiden \mathbf{p} ja \mathbf{s}_i keskinäistä
sijaintia (niiden longituedeja kulmakoordinaattimuodossa). Tulokse-
na on vektorin \mathbf{v}' kulma vektorista \mathbf{o} positiiviseen kiertosuuntaan
vektorin \mathbf{p} ympäri.
- (c) Lasketaan reaaliarvoinen lokero x suoraan kulmasta ψ yhtälöllä
 $x = n\psi/2\pi$ ja sen murto-osa $q = x - \lfloor x \rfloor$. Mikäli asetuksissa on
määriteltä, että käytetään kokonaislukulokeroita, pyöristetään q ko-
konaisluvuksi 0 tai 1.
- (d) Signaaliarvoa lokerossa $\lfloor x \rfloor$ lisätään arvolla $(1-q)d_i$ ja signaaliarvoa
lokerossa $\lfloor x \rfloor + 1 \pmod n$ lisätään arvolla qd_i . Näiden lokeroitten
painoja lisätään vastaavasti arvoilla $1 - q$ ja q .

Jokaisen lokeron sijainti lasketaan pyörittämällä vektoria \mathbf{o}'' vektorin \mathbf{p}
ympäri $2\pi/n$ lisäyksin. Vektori \mathbf{o}'' on skaalattu vektorista \mathbf{o} siten, että
 $\mathbf{p}' + \mathbf{o}''$ on yksikkövektori eli rajoittamalla vektorin \mathbf{o} pituus yksikköpal-
lolle sen alkaessa pisteestä \mathbf{p}' : $\mathbf{o}'' = (1 - |\mathbf{p}'|^2)\mathbf{o}$. Rotaatio tuottaa vektorin
 \mathbf{v}' , josta saadaan $\mathbf{s}_i = \mathbf{p}' + \mathbf{v}'$ joka puolestaan muunnetaan suuntausmuo-
toon \vec{s}_i .

Normalisaatioalgoritmi tuottaa *painotettuja* signaaliarvoja, jotka muunne-
taan varsinaisiksi signaaliarvoiksi jakamalla jokainen arvo vastaavalla painolla.
Painotettu muoto on hyödyllinen silloin, kun rengas keskiarvoistetaan samas-
sa vastaanotinryhmässä olevilta vastaanottimilta peräisin olevien renkaiden
kanssa, jolloin sekä signaaliarvot että painot voidaan yksinkertaisesti laskea
yhteen.²

6.2 Signaalin kohinan vaimentaminen

Simulaation oletusarvojen perusteella yhdessä näytteistysrenkaassa olevan ko-
hinan amplitudi on suuruusluokkaa 5 Janskya. Koska useimmat pistelähteet
ovat kirkkaudeltaan pienempiä, näin kohinainen signaali olisi täysin hyödy-
tön. Sen takia Planckin kuvausstrategia kuvaakin samaa taivasympyrää 60
kierrosta, jolloin keskiarvoistamalla nämä näytteistysrenkaat saadaan kohinaa
yhdessä renkaassa vähennettyä kertoimella $\sqrt{60}$ alle yhden Jansdyn tasolle.

QDS tekee edellä mainitun keskiarvoistuksen joka tapauksessa. Sen lisäksi
kohinaa voidaan vielä vaimentaa jatkamalla keskiarvoistusta kahdella tasolla,
jotka ovat valinnaisia ja konfiguroitavissa erikseen QDS:ssä.

²Tulos painottuu luonnollisesti niitä vastaanottimia kohti, joiden alkuperäisessä datassa
oli eniten näytteitä.

Ensinnäkin voidaan ryhmitellä yhteen samalla taivasympyrällä olevat vastaanottimet, joilla on samat ominaisuudet (taajuus ja keilanleveys), ja keskiarvoistaa näiden signaalit. Koska Planckin LFI-vastaanotintorvissa on jokaisessa kaksi lähinnä polarisaatioltaan eroavaa vastaanotinta³, saadaan yhteen ryhmään ainakin kaksi vastaanotinta. Lisäksi kuvasta 2.3 nähdään, että useimmiten samalla taivasympyrällä on kaksi syöttötorvea, jolloin saadaan neljän vastaanottimen ryhmiä. Planckin LFI-datalle siis saadaan yksi ryhmä taajuudella 30 GHz, kaksi ryhmää taajuudella 44 GHz ja kolme ryhmää taajuudella 70 GHz.

Vastaanotinryhmittelyn lisäksi voidaan ottaa käsiteltävä rengas ja sitä ympäröivät renkaat säteellä r (r välittömästi edeltävää ja r välittömästi seuraavaa rengasta) ja keskiarvoistaa nämä yhteen. Kohinanvaimennuskerroin on tällöin $\sqrt{2r+1}$. Haittapuolena tässä menetelmässä on se, että pistelähteiden vuoarvot ovat niiden oikeita arvoja pienemmät, koska huipun vuo onkin keskiarvo maksimista ja sen ympärillä olevista arvoista. Myös paikkatarkkuus voi teoriassa kärsiä jonkin verran. Tämän menetelmän näkökulmasta tärkeää on se, että renkaan näytteet ovat lähellä vastaavia näytteitä muissa renkaissa. Tämä tarkoittaa sitä, että vaikka Planckin renkaat menevätkin ristiin lähellä napoja, voidaan niitä keskiarvoistaa samalla tavalla kuin jos renkaat olisivat peräkkäin lieriöpinnalla.

Kuvassa 6.2 on demonstroitu eri kohinanvaimennustasot melko tasaisella pätkällä Planckin 30 GHz:n signaalia. Tästä huomataan, että yhdistämällä molemmat lisävaimennusmenetelmät (kuva 6.2(c), $r = 4$) päästään jo aivan riittävän lähelle kohinatonta signaalia (kuva 6.2(d)). Tällöin vaimennuskerroin verrattuna yhden vastaanottimen signaaliin on $\sqrt{4 \cdot 9} = 6$.

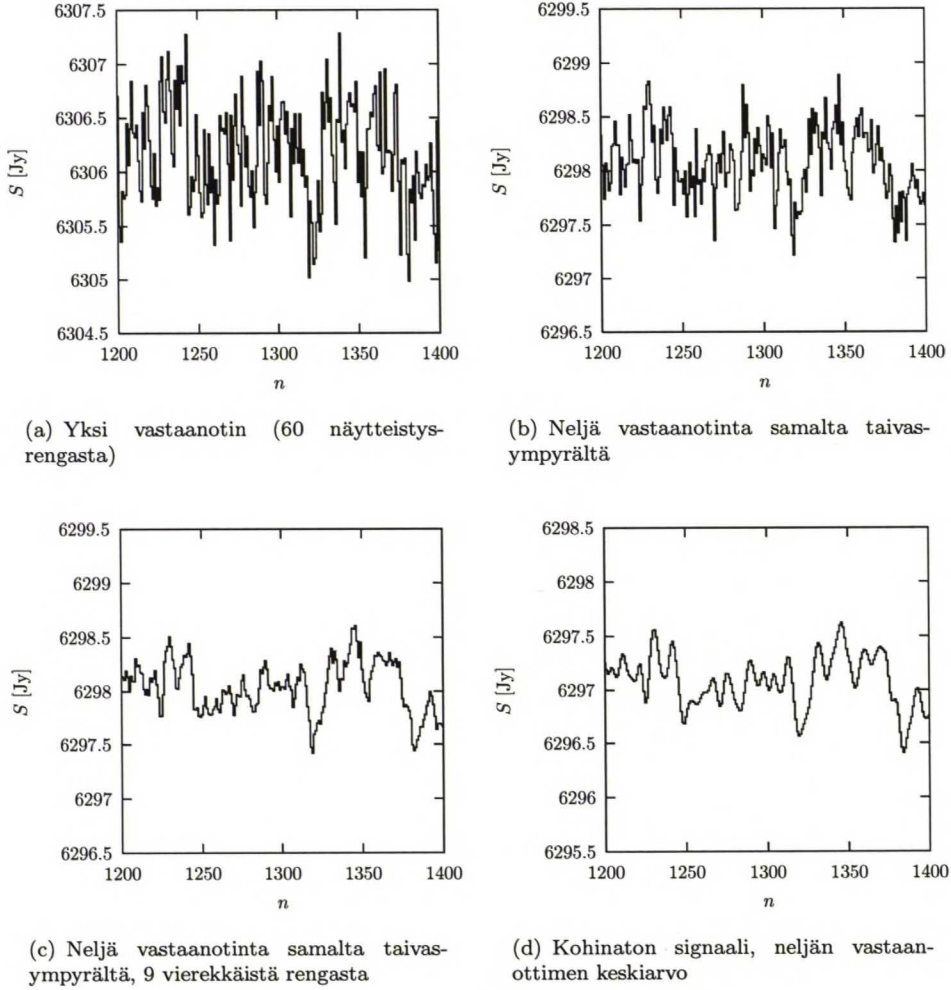
Täysin kohinattomassakin signaalissa on taustasäteilystä, Linnuradasta, kaasupilvistä yms. johtuvia epätasaisuuksia. Nämäkin ovat pistelähteitä etsivän QDS:n kannalta kohinaa, mutta niitä ei voi poistaa keskiarvoistamalla.

Planckin signaalissa on mukana instrumenttien valkoista ja $1/f$ -kohinaa⁴ sekä satelliitin jäähdytysjärjestelmän kohinaa. Kaksi jälkimmäistä sisältävät hyvinkin suuria pitkäaaltoisia komponentteja, mutta tällaiset häviävät melko hyvin keskiarvoistuksessa ja myöhemmin suotimien kaistanpäästöominaisuuksien ansiosta (ks. kuitenkin kohta 6.4.2). Esimerkiksi $1/f$ -kohinan vaikutus keskiarvoistuksen jälkeen on lähinnä vakiolisäys renkaan signaalitasoon [Ma2002, Ja1996]. QDS:n kannalta kohinan voidaan siis katsoa olevan lähinnä valkoista.

Koska suotimet karsivat pois myös korkeat taajuudet, ei vierekkäisten renkaiden keskiarvoistaminen ole välttämättä niin hyödyllistä kuin mitä sen vaikutus suodattomattomana antaisi olettaa. Kannattaa siis olla varovainen r :n suuruuden valinnassa, sillä jo arvolla $r = 1$ voi olla mahdollista saada aivan yhtä hyviä lopullisia tuloksia kuin kuvassa 6.2 käytetyllä $r = 4$. Ja

³Keilan muodot eroavat myös jonkin verran, mutta eivät merkittävästi QDS:n kannalta.

⁴Myös vaaleanpunainen kohina (*pink noise*) tai *flicker noise*.



Kuva 6.2: Sama osa signaalista eri keskiarvoistusvaiheiden jälkeen, sekä vertailukohtana kohinaton signaali. (Huom: signaalin keskitasot vaihtelevat pitkäaaltoisen kohinan takia.)

koska varsinainen tarkoitus on pienentää kohinan amplitudia jotta vältetään valheelliset pistelähdehavainnot, ei enemmästä keskiarvoistamisesta ole enää hyötyä sen jälkeen, kun varsinaisen kohinan amplitudi on laskenut signaalin taustan vaihtelun amplitudin alle.

6.3 Signaaliarvojen muunto vuontiheydeksi

Rengasdatan lukemisen ja normalisoinnin jälkeen antennilämpötilana T_a olevat signaaliarvot muunnetaan vuontiheydeksi S Janskyina⁵. Ohjelmassa on

⁵1 Jy = 10^{-26} W/m²Hz

kolme tapaa tehdä tämä lineaarinen muunnos, riippuen mitä asetusmuuttujia on määritelty. Tavat ovat seuraavat, joista ensimmäistä mahdollista käytetään:

1. Mikäli suora muuntokerroin J on määritelty, käytetään yksinkertaista kaavaa

$$S = JT_a \quad (6.1)$$

2. Mikäli antennin efektiivinen pinta-ala A_e on määritelty, käytetään kaavaa

$$S = C \frac{2k}{A_e} T_a \quad (6.2)$$

3. Muussa tapauksessa efektiivistä pinta-alaa arvioidaan kaavalla $\Omega A_e = \lambda^2$, missä taas parametria Ω arvioidaan kaavalla $\Omega = 1.13\omega^2$. Tästä saadaan muunnoskaava

$$S = C \frac{2.26k\omega^2\nu^2}{c^2} T_a \quad (6.3)$$

Kaksi viimeistä tapaa ovat hyödyllisiä lähinnä silloin, kun kalibrointia tunnettua lähdettä vasten ei voida tehdä. Mikäli kalibrointi voidaan tehdä, tulisi käyttää ensimmäistä tapaa.

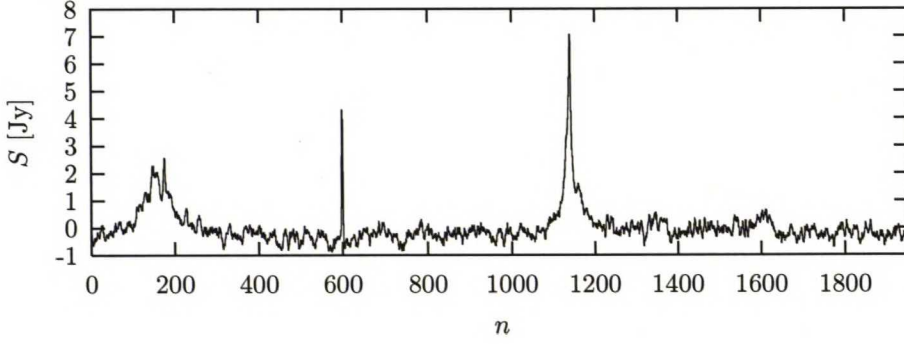
Seuraavassa on listattu käytetyt parametrit ja niiden selitykset:

J	Suora muuntokerroin.
C	Korjauskerroin kaavan hienosäätöä varten.
k	Boltzmannin vakio.
A_e	Vastaanottimen antennin (keilan) efektiivinen pinta-ala (m^2).
Ω	Vastaanottimen keilan koko (steradiaaneja).
ω	Vastaanottimen keilan puoliarvoveveys (FWHM) (radiaaneja).
λ	Vastaanottimen nimellinen aallonpituus, $\lambda = c/\nu$.
ν	Vastaanottimen nimellinen taajuus.
c	Valon nopeus tyhjiössä.

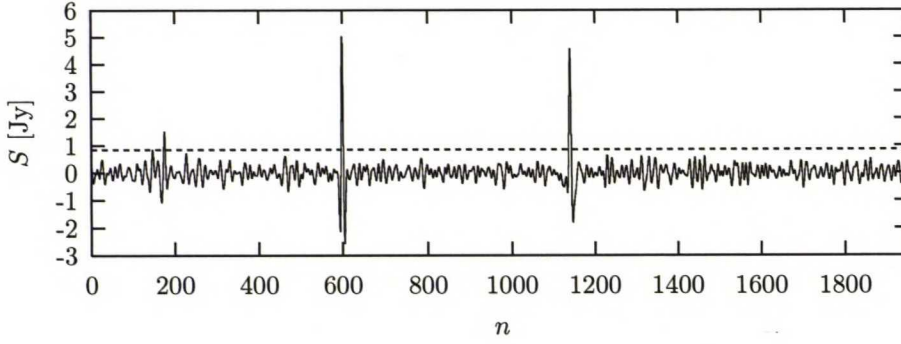
6.4 Suodatus

QDS käyttää pistelähteiden etsimiseen Meksikolaishattusuodinta (Mexican hat wavelet filter). Sen määritelmä on annettu kappaleessa 4.1.2.

Kuvassa 6.3 on simulaattorin tuottama esimerkkirengas 30 GHz:n vastaanottimilla. Kuva havainnollistaa hyvin, miten suodin poimii pistelähteet esiin signaalista. Kohdassa $n = 600$ on 5 Jy:n vahvuinen pistelähde, kaikki muu signaali on peräisin Linnunradasta yms. taustasta. Erityisen huomionarvoinen on Linnunradan kirkas kohta kohdassa $n \approx 1150$, koska tarpeeksi kapeana se jää näkyviin pistelähteenä suodatuksen jälkeen. Tällaiset Linnunradasta johuvat väärät hälytykset pitää ottaa jollain tavalla huomioon. Yksi vaihtoehto



(a) Suodattamaton signaali (keskitaso siirretty)

(b) Suodatettu signaali, kynnyсарvo $4\sigma \approx 0,85$ Jy merkitty katkoviivalla.**Kuva 6.3:** Esimerkkirengas, jossa 5 Jy pistelähde kohdassa $n = 600$

on tallentaa tunnetut sellaiset QDS:n tietokantaan, mutta jo pelkkä havainnon galaktisten koordinaattien esittäminen auttaa paljon: mikäli galaktinen latitudi on kovin lähellä nollaa, on havainto suuremmalla todennäköisyydellä peräisin Linnunradasta kuin jos se olisi lähempänä galaktisia napoja.

Suodatetusta signaalista näkee myös meksikolaishattusuotimelle tunnusomaiset negatiiviset sivupiikit pistelähteen kohdalla. Suodatetun signaalin kuvaan on myös merkitty kynnyсарaja 4σ , joka on laskettu kappaleessa 6.5 esitellyllä tavalla.

QDS luo suotimen aika-avaruudessa ja muuntaa sen sitten Fourier-muunnoksella Fourier-avaruuteen. Suodatusoperaatio ensin muuntaa suodatettavan signaalin Fourier-muotoon, kertoo sen sitten suotimella ja lopuksi tekee käänteisen Fourier-muunnoksen tulokselle. Tämä on tehokas menettelytapa, sillä FFTW-kirjasto suorittaa kaikki muunnokset aikakompleksisuudella $O(N \log N)$. Suotimen konvoluution laskemisen aikakompleksisuus ilman Fourier-muunnosta olisi $O(N^2)$.

6.4.1 Suotimen vahvistuksen kompensointi

Kaistanpäästösuoitimenä meksikolaishattusuodin vahvistaa signaalin heilahte-
luja eri kertoimella riippuen niiden leveydestä. Jotta havaitun pistelähteen
vuo voitaisiin suodatuksen jälkeen määrittää, pitää tietää suotimen vahvis-
tus (*gain*) pistelähteille. Signaalin pistelähteet ovat gaussisia huippuja, joilla
on vastaanottimen keilan puoliarvoleveys ω . Ideaalista pistelähdettä kohdassa
 $t = 0$ ilman taustaa kuvaa siis funktio

$$p(t) = e^{-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2}, \quad (6.4)$$

missä leveysparametri σ saadaan yhtälöstä

$$\sigma = \frac{\omega}{2\sqrt{2\ln 2}}. \quad (6.5)$$

Haettu vahvistus on suodatetun pistelähteen maksimiarvo (myöskin koh-
dassa $t = 0$) suhteessa suodattamattoman signaalin maksimiarvoon $p(0) = 1$.
Jos suotimen leveysparametri siis on R , saadaan vahvistukselle G kaava

$$G(R, \omega) = \int_{-\infty}^{\infty} p(t)f(t)dt = \int_{-\infty}^{\infty} e^{-4\ln 2\left(\frac{t}{\omega}\right)^2} \left(1 - \left(\frac{t}{R}\right)^2\right) e^{-\frac{1}{2}\left(\frac{t}{R}\right)^2} dt \quad (6.6)$$

Integraalin laskeminen ja sievennys antaa yhtälön

$$G(R, \omega) = \omega \frac{\sqrt{2\pi} \ln 256}{\left(\left(\frac{\omega}{R}\right)^2 + \ln 256\right)^{3/2}} \quad (6.7)$$

Jos suotimen leveys ilmaistaan keilan puoliarvoleveyden moninkertana $R = k\omega$, mikä käytännössä on ainoa järkevä tapa, pätee vahvistukselle $G(k, \omega) = \omega G(k, 1)$.

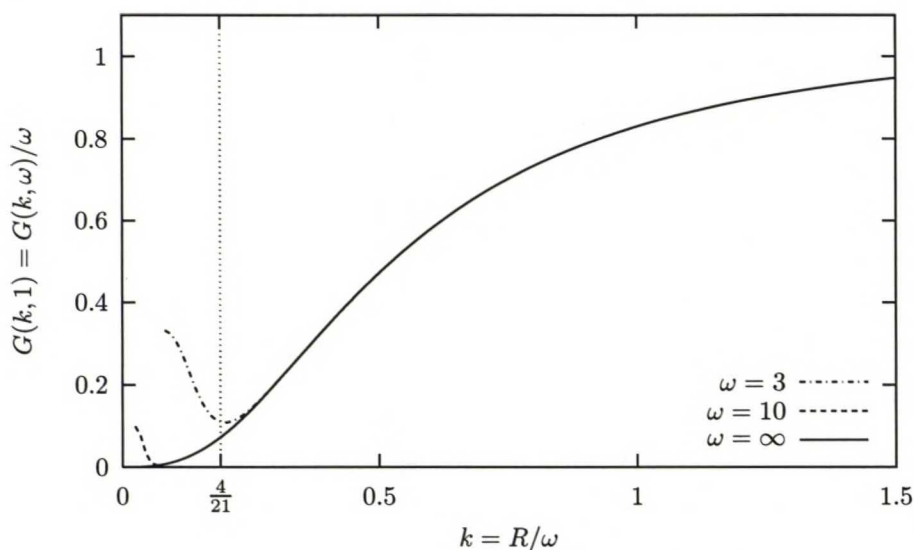
Käytännön tapauksessa signaali on diskreetti, jolloin integraali muuttuu summaksi ja se voidaan laskea numeerisesti. Tarkastelemalla kuvaa 4.1 voi-
daan todeta, että väli $[-6R, 6R]$ on käytännössä riittävän tarkka integrointi-
summan laskemiseen. Tämän perusteella saadaan vahvistuksen numeeriseksi
laskukaavaksi

$$G'(k, \omega) = \sum_{n=-6k\omega}^{6k\omega} e^{-4\ln 2\left(\frac{n}{\omega}\right)^2} \left(1 - \left(\frac{n}{k\omega}\right)^2\right) e^{-\frac{1}{2}\left(\frac{n}{k\omega}\right)^2} \quad (6.8)$$

Jos käytetään normalisoitua muotoa $G'(k, \omega)/\omega = G'(k, 1)$ ja annetaan puo-
liarvoleveyden kasvaa äärettömäksi, päädytään takaisin kaavaa 6.7 vastaavaan
integraalimuotoon, eli toisin sanoen $\lim_{\omega \rightarrow \infty} G'(k, \omega)/\omega = G(k, 1)$.

Kaavassa 6.8 n kuvaa näytteitä. Taulukosta 2.1 nähdään, että Planckin
LFI-vastaanottimien keilan puoliarvoleveys on noin kolme näytettä, jolloin

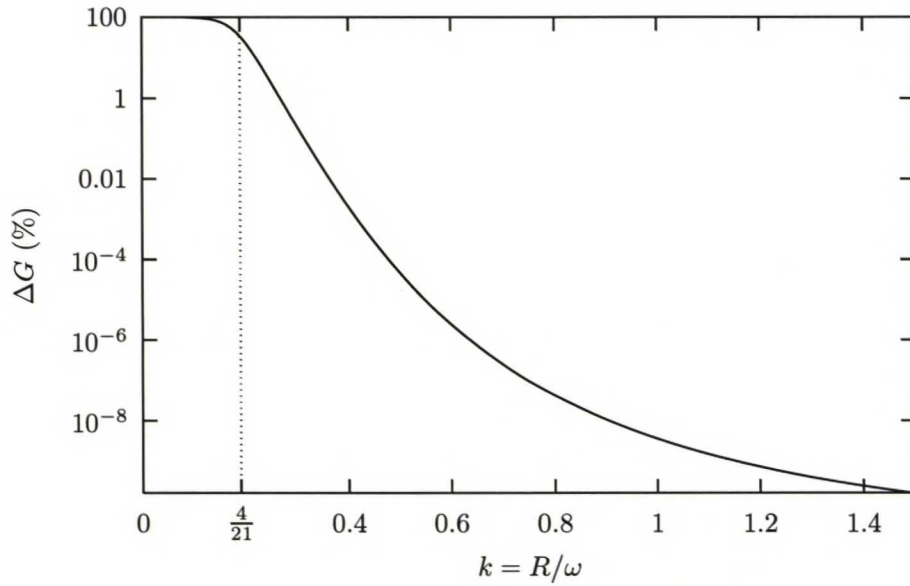
kaavaan 6.8 asetettaisiin $\omega = 3$. Diskreetissä tapauksessa suotimen leveys ei voi olla kovin pieni, koska muuten näytteiden rajallisuus muuttaa sen muotoa merkittävästi. Suodin pelkistyy yksikkönäytefunktiksi suunnilleen kohdassa $4R = 1 \iff k = 1/(4\omega)$, jolloin siis vahvistuskertoimeksi tulee yksi. Toisaalta pienimmäksi mielekkääksi meksikolaishattusuotimeksi voidaan katsoa sellainen, jossa on päähuipun vieressä molemmat negatiiviset sivuhuiput suunnilleen oikean kokoisena. Koska sivuhuipun paikat ovat suunnilleen kohdassa $R = \pm 1,75$, on mielekkäällä suotimilla siis $1,75R > 1 \iff k > 4/(7\omega)$. Nämä ilmiöt näkyvät kuvassa 6.4, johon on kaavan 6.7 kuvaajan lisäksi piirretty numeerisesti lasketut vahvistuskäyrät arvoilla $\omega = 3$ ja $\omega = 10$. Normalisoinnin vuoksi näytefunktiksi pelkistyminen tuottaa vahvistuskertoimet $\frac{1}{3}$ ja $\frac{1}{10}$. Kuvassa on myös merkitty mielekkään suotimen leveyden raja-arvo $\frac{4}{21}$ tapauksessa $\omega = 3$.



Kuva 6.4: Meksikolaishattusuotimen vahvistus pistelähteille sen keilanleveyteen nähden suhteellisen leveyden R/ω funktiona, normalisoituna tapaukseen $\omega = 1$.

Kaavan 6.7 käyttö diskreetissä tapauksessa tuottaa siis virheen vahvistuskertoimeen. Kuvasta 6.5 nähdään, että Planck LFI:n kannalta oleellisessa tapauksessa $\omega \approx 3$ virhe on aivan riittävän pieni, mikäli $k > 0,3 \iff R > 0,3\omega$. Tämän voidaan olettaa toteutuvan — vaikkakin juuri ja juuri — optimaalisen suotimen leveyden tapauksessa artikkelin [Vi2001] perusteella.

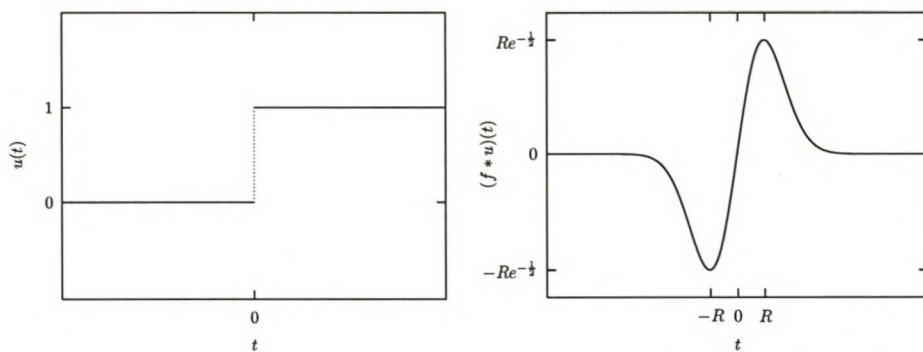
Vaihtoehtoinen tapa laskea suotimen vahvistuskerroin on taulukoida kaavan 6.8 arvoja käytetyllä keilanleveydellä ja käyttämällä väliarvoille (lineaarista) interpolointia. Tämä on kuitenkin järkevää ainoastaan silloin, kun pistelähteen muoto poikkeaa *merkittävästi* gaussisesta (kaava 6.4), jolloin koko meksikolaishattusuotimen käyttö ei ole välttämättä järkevää.



Kuva 6.5: Suhteellinen virhe käytettäessä kaavaa 6.7 diskreetissä tapauksessa $\omega = 3$.

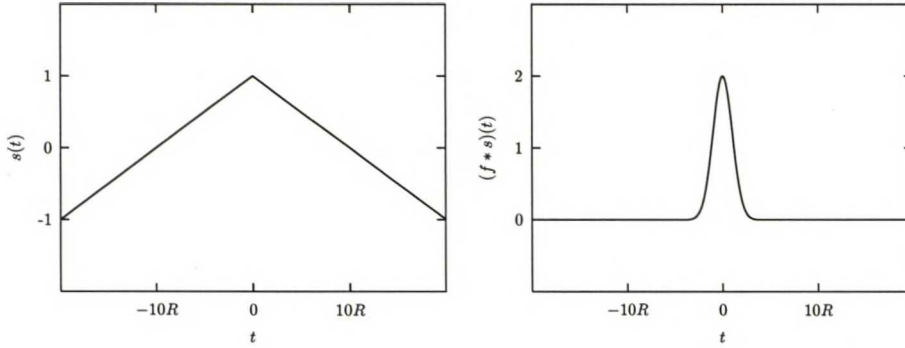
6.4.2 Epäjatkuvuudet suodatettavassa signaalissa

Huippujen etsimisen kannalta (ja muutenkin) epäjatkuvuuskohdat suodatettavassa signaalissa ovat ongelmallisia. Tätä havainnollistaa hyvin kuvassa 6.6 esitetty yksikköaskelfunktion suodatus: epäjatkuvuuskohdasta etäisyydellä R on $Re^{-\frac{1}{2}}$ korkeat piikit. Tällaisten efektien syntyminen ei vaadi varsinaisesti epäjatkuvuutta, tarpeeksi jyrkkä nousu signaalissa riittää. Myös signaalin derivaatan epäjatkuvuuskohta aiheuttaa piikin suodatettuun signaaliin (kuva 6.7).



Kuva 6.6: Meksikolaishattusuotimen vaikutus yksikköaskelfunktion.

Oikean signaalin tapauksessa ei voida tarkasti ottaen puhua jatkuvuudesta, koska signaali on ensinnäkin diskreetti ja toisekseen se on kohinainen, ja



Kuva 6.7: Meksikolaishattusuotimen vaikutus derivaatan epäjatkuvuuskohtaan.

kohinan mielletään olevan osa “jatkuvaa” signaalia. Määritelläänkin siis *kohina-jatkuvuuden* tarkoittavan sitä, että signaalissa mahdollisesti olevia epäjatkuvuuskohtia ei voi erottaa kohinasta. Formaalisti tämän voisi ilmaista siten, että epäjatkuvuudesta aiheutuvien piikkien suuruus ($Re^{-\frac{1}{2}}$) on oltava pienempi kuin signaalin kohinataso σ . Käytännössä tällainen vaatimus on kuitenkin mahdoton todentaa. Samoin kohinaiselle signaalille on mahdotonta johtaa järkeviä derivaatan arvoja, joten derivaatan epäjatkuvuuksille ei voi juurikaan tehdä mitään analyttisesti. Käytännössä derivaatan epäjatkuvuuden aiheuttaman “kulman” terävyys on usein tarpeeksi pieni kohinaan verrattuna, jolloin siitä aiheutuva efekti ei erotu kohinasta.

Planckin raakasignaali on kohinajatkuvaa, mikäli siinä ei ole toimintavirheestä tai muusta syystä aiheutuvia aukkoja. Siitä otettujen pätkien keskiarvo (kappaleessa 6.1 määritelty normalisointi) on myöskin kohinajatkuva. Koska kyseessä on taivaalta kuvattu rengas, on signaali yhden renkaan sisällä myös *syklisesti* kohinajatkuva *mikäli* myös siinä oleva kohina on syklisesti kohinajatkuvaa. Planck-simulaattorin perusteella satelliitin jäähdytysjärjestelmästä aiheutuva häiriö on hyvin pitkäaaltoista ja suuriamplitudista⁶, joten se todennäköisesti aiheuttaa renkasiin syklisen kohinaepäjatkuvuuden. Koska suodatusoperaatio on syklinen konvoluutio, pitää nämä epäjatkuvuudet poistaa renkaista ennen suodatusta väärin pistelähdehavaintojen ehkäisemiseksi. Tämä toivottavasti tehdään Planckin datankäsittelyputkessa jo ennen QDS:ää, mutta kaiken varalta QDS:n täytyy varautua tekemään se itse.

Syklisen kohinaepäjatkuvuuden aiheuttamien efektien eliminointiin on kaksi lähestymistapaa. Ensimmäinen on jatkaa signaalia kummastakin päästä kyseisen pään signaalitason määrittämällä vakiosignaalilla tarpeeksi pitkälle (noin $5R$ etäisyydelle), jolloin epäjatkuvuudesta aiheutuvat piikit jäävät

⁶Signaali mitataan vastaanottimen lämpötilasta, ja jäähdytin ei pysty pitämään instrumenttien lämpötilaa vakiona vaan siinä esiintyy oskillaatiota.

varsinaisen signaalin ulkopuolelle. Efektiivisesti tämä tekee suodatusoperaatiosta ei-syklisen konvoluution. Toinen tapa on tehdä signaalista kohinajatkuvaa määrittämällä kummankin pään signaalitasot ja vähentämällä signaalista näiden pisteiden määrittämä suora. Tällöin signaalitasot siirtyvät kummassakin päässä noin nollaan ja signaali on kohinajatkuvaa.

Jälkimmäisessä tavassa jokainen näyte käydään lineaarisesti läpi, ensimmäisessä taas näytteiden määrää lisätään, mikä vaikuttaa jonkin verran suodatuksen vaatimaan aikaan. Kummassakin tavassa päiden signaalitasot määritettäisiin käytännössä ottamalla esimerkiksi kymmenen näytteen keskiarvo.

Ensimmäisen tavan etu on sen suhteellinen yksinkertaisuus ja tehokkuus, koska syklisyysominaisuuden vuoksi kaikki lisäykset voidaan tehdä signaalin perään, sekä se, että se ei muuta varsinaista signaalia. Tätä tapaa ei kuitenkaan voida käyttää mikäli signaalin alkukohtaa on tarpeen siirtää, koska silloin kohinaepäjatkuvuuskohta ei ole enää datarenkaan lopussa. Alkukohdan siirtäminen on kuitenkin normalisointivaiheessa tarpeen, koska tarkkaan ottaen peräkkäisten renkaiden ei voida olettaa alkavan samasta kohtaa. Vielä tärkeämpi syy on se, että eri vastaanottimien renkaat alkavat eri kohdista, ja samaa taivasympyrää kuvaavien vastaanottimien dataa halutaan ehdottomasti keskiarvoistaa kohinan vaimentamiseksi.

Toisessa lähestymistavassa tätä ongelmaa ei ole, koska signaali pakotetaan kohinajatkuvaksi. Tämäkin menetelmä on yksinkertainen ja tarpeeksi tehokas ($O(N)$). Varjopuolena on signaalin muokkaaminen, jonka tosin ei käytännössä *pitäisi* vaikuttaa merkittävästi suodatuksen tulokseen, sekä mahdollisuus, että kohinaepäjatkuvuuskohtaan jää operaation jälkeen niin voimakas "kulma", että se erottuu suodatuksen jälkeen kohinasta.

Koska konvoluutio (ja Fourier-muunnos) on lineaarimuunnos, voi suodatuksen periaatteessa tehdä joko ennen tai jälkeen keskiarvoistuksen ja päätyä samaan tulokseen. Mikäli suodatus tehdään ensin, on signaalin jatkaminenkin vartenotettava vaihtoehto. Esimerkiksi jos signaali suodatetaan jo ennen normalisointia, täytyy käsitellä vain yksi epäjatkuvuuskohta ja suodatetussa signaalissa niitä ei ole, joten renkaiden alkukohdan siirto voidaan tehdä. Koska suodatuksen aikakompleksisuus kuitenkin on $O(N \log N)$ ja keskiarvoistuksen $O(N)$, kannattaa keskiarvoistus tehdä ensin jolloin suodatettavana on huomattavasti vähemmän näytteitä. Lisäksi jotkin operaatiot tarvitsevat keskiarvoistetun suodattamattoman signaalin.

Näillä perusteilla QDS käyttää signaalitasoa siirtävää menetelmää kohinaepäjatkuvuuskohtien käsittelyyn. Käsittely voidaan myös kytkeä pois päältä, mikäli käsiteltävä data tiedetään syklisesti kohinajatkuvaksi renkaittain.

6.5 Signaalin kohinatason määrittäminen ja kynnysarvoistus

Vain tietyn kynnysarvon ylittävät huiput suodatetussa signaalissa katsotaan pistelähdelöydöiksi. Tämä kynnysarvo on mahdollista määritellä käsin Janskyina, mutta näin on yleensä järkevää tehdä ainoastaan interaktiivisen käyttöliittymän kautta. Paljon parempi tapa on käyttää automaattista kynnysarvon määrittystä, jossa kynnysarvo määritellään signaalin kohinatason moninkertana. Kohinatasoksi on määritelty signaalin RMS-kohina eli signaalissa olevan kohinan keskihajonta σ . Kynnysarvon sekä vuon virheen määrittämisen kannalta kohinaa ovat kaikki ne kohdat signaalista, jotka eivät ole pistelähteitä. Tavoitteena on siis saada signaalin kohinatasoon mukautuva kynnysarvo.

Mutta kuinka selvittää signaalin kohinataso? Ongelma on siinä, että haluttu arvo on *kohinan* keskihajonta, mutta (suodatetussa) signaalissa on mukana kohinan lisäksi pistelähteitä, ja keskihajonnan laskeminen koko signaalista sisältää myös pistelähdehuippujen vaikutuksen. Tällä tavalla saatu keskihajonta saattaa olla liian suuri ollakseen sopiva kuvaamaan kohinatasoa. Tämä vaikutus on erityisen merkityksellinen, mikäli signaalissa on hyvin kirkkaita pistelähteitä kuten planeettoja, joiden kirkkaus on jopa tuhansien Janskyjen luokkaa.

QDS:ssä tämän ongelman ratkaisemiseksi käytetään menetelmää, jossa signaali rajoitetaan nollan ympärillä⁷ välille $[-L, L]$ ja keskihajonta σ lasketaan tästä rajoitetusta signaalista. Menetelmä perustuu siihen, että pistelähdenäytteiden osuus koko renkaan näytteistä on pieni (normaalisti paljon alle 10%), joten isoimpien huippujen poistamisen pitäisi kuvata signaalin kohinaa riittävästi hyvin. Raja L määritellään siten, että N prosenttia signaalin näytteistä ovat sitä pienempiä, $N \in [50, 100]$, ja se saadaan helposti järjestämällä renkaan näytteet suuruusjärjestykseen ja ottamalla näytteen $nN/100$ arvo, missä n on näytteiden määrä renkaassa. Tällä tavalla saatu σ kerrotaan vielä korjauskerroinella C , jolla kompensoidaan rajoituksen aiheuttamaa virhettä. Kohinaa voidaan siis arvioida kaavalla

$$\sigma = C\sigma_N \quad (6.9)$$

missä σ_N on kohinataso laskettuna siten, että rajoitus on N prosentin kohdalla. Vaikka meksikolaishattusuodin tuottaa pistelähteen molemmille puolille negatiiviset huiput, tässä menetelmässä käytetään silti näytteiden varsinaisia arvoja eikä niiden itseisarvoja määritettäessä rajoitustasoa L .

Tällä menetelmällä voimakkaimpien pistelähteiden vaikutus kohinatasoon saadaan eliminoitua. Arvojen $C = 1$ ja $N = 100$ käyttäminen on efektiivisesti sama kuin jos käytettäisiin koko signaalin keskihajontaa. Muuttujat C ja N voidaan määritellä QDS:n asetustiedostossa.

⁷Suodatus poistaa tasavirtakomponentin, joten signaalin keskitaso on 0.

Seuraavissa kappaleissa on kuvattu testi, jolla pyrittiin selvittämään, onko edellä kuvattu menetelmä Planckin tapauksessa yleensäkin kelvollinen kohinatason määrittämiseen, ja mitkä ovat parametrien C ja N optimaaliset arvot. Testissä simuloitiin erilaisia Planckin datarenkaita ja tulokset analysoitiin.

6.5.1 Testidata

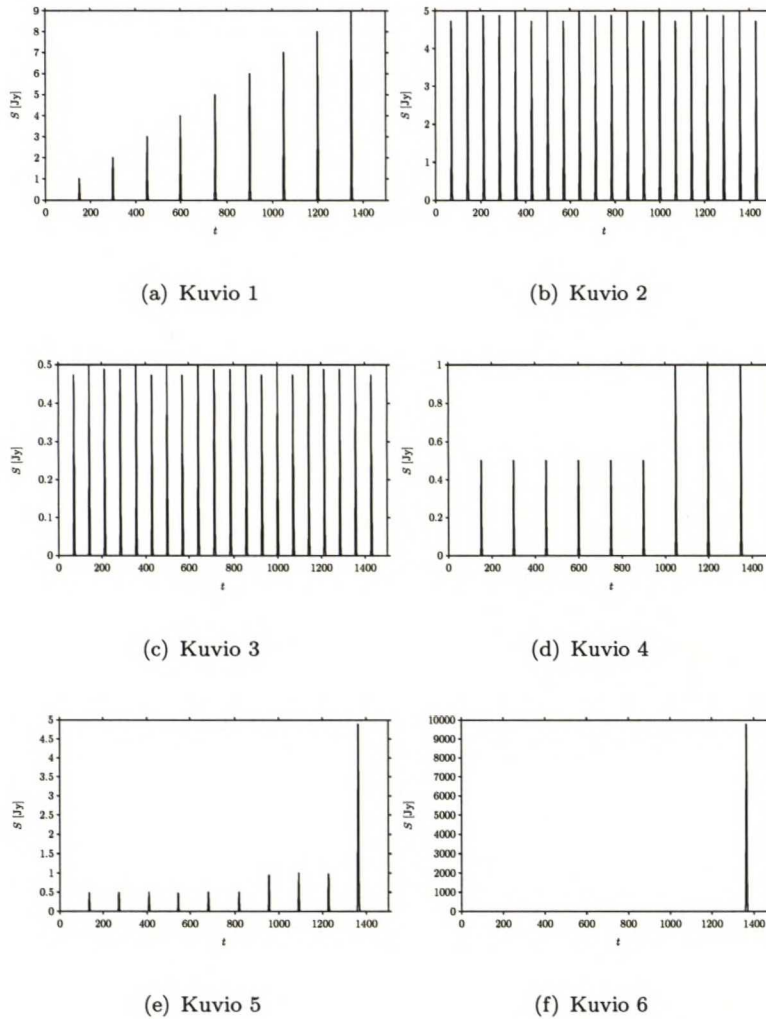
Menetelmän hyöty olisi siinä, että se on suhteellisen yksinkertainen eikä se tarvitse muuta dataa kuin itse renkaan. Planckin signaalidatan *tyyppi* ei kuitenkaan ole täysin tuntematon, vaan joitakin sen ominaisuuksia tiedetään etukäteen. Ensinnäkin se on suodatettu tietynlevyisellä meksikolaishattusuotimella. Toiseksi ennen suodatusta data koostuu lähinnä valkoisesta kohinasta, pistelähteistä tietynlevyisinä gaussisina huippuina sekä suurista Linnunradasta ja muista lähteistä peräisin olevista rakenteista jotka suodatus pääosin poistaa. Tässä testissä käytetty simulaatiodata koostuu valkoisesta ja simuloidusta kohinasta sekä pistelähteiden gaussisista huipuista.

Testiä varten luotiin kaksi kohinataustaa A ja B Perl-lausekkeella "`rand(2) - 1`", joka tuottaa pseudosatunnaislukuja välillä $[-1, 1]$. Lisäksi Planck-simulaattorilla luotiin kaksi simuloitua kohinataustaa C ja D simulaattorin kohinamoduulin ja jäädytinkohinamoduulin oletusasetuksilla ilman minkäänlaista signaalia. Simulaattorilla simuloitiin yksi tunti, jonka 60 näyterengasta keskiarvoistettiin yhteen.

Planck-simulaattorin tuottama signaali QDS:n käsittelyn jälkeen on kohinavaimennettua, joten siinä oleva kohina ei ole yhtä korkeataajuisia kuin puhdas valkoinen kohina (ks. kuva 6.2) taustasta johtuvan häiriön ollessa suurempi tekijä. Tämän huomioon ottamiseksi ja testidatan lisäämiseksi luotiin kohinataustat $A' \dots D'$ suodattamalla taustat $A \dots D$ viiden levyisellä liukuvala keskiarvoistussuotimella. Suotimen leveys arvioitiin silmämääräisesti siten, että saadut kohinataustat muistuttavat paremmin simulaattorin tuottamissa renkaissa olevaa kohinaa. Kohinan amplitudi on noin välillä 0,5 ja 1 Janskya simulaattoriin perustuen, joten tässä mielessä kohinataustojen pitäisi olla realistisia.

Kuten taulukosta 2.1 nähdään, Planckin LFI-vastaanottimien keilojen puoliarvovaleveydet ovat kaikissa tapauksissa noin kolme näytettä, joten tätä arvoa käytettiin luotaessa pistelähdehuippuja kaavojen 6.4 ja 6.5 perusteella. Eri vahvuiset pistelähteet saadaan kertomalla kaava 6.4 pistelähteen vuolla S . Testiä varten luotiin kuusi pistelähdehuippukuviota (kuvassa 6.8), joissa huiput ovat renkaalla tasavälein:

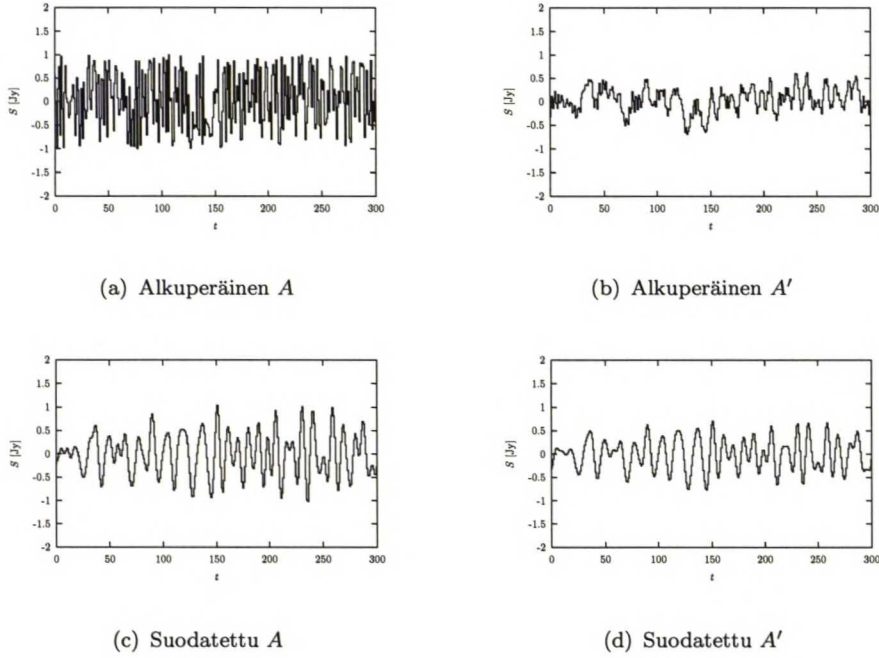
1. Testikuvio: yhdeksän lähdettä kirkkaudeltaan 1–9 Jy.
2. Kirkkaat huiput: 20 lähdettä kirkkaudeltaan 5 Jy.
3. Himmeät huiput: 20 lähdettä kirkkaudeltaan 0,5 Jy.



Kuva 6.8: Testissä käytetyt huippukuviot

4. Tyypillinen: kuusi 0,5 Jy lähdetä ja kolme 1 Jy lähdetä.
5. Tyypillinen + linnurata: sama kuin tyypillinen ja lisäksi yksi 5 Jy:n lähde. (Simulaattorissa tällainen huippu tulee Linnunradasta ainakin yhdessä kohtaa, ks. kuva 6.3.)
6. Tyypillinen + planeetta: sama kuin tyypillinen ja lisäksi yksi 10000 Jy:n lähde. Simuloi tapausta, jossa renkaassa on hyvin kirkas planeetta (esim. Jupiter).

Jokainen näistä huippukuviosta lisättiin jokaiseen kohinataustaan, jolloin saatiin 48 testirengasta, jotka suodatettiin meksikolaishattusuotimella leveysarvolla $R = 3$. Suotimen leveydeksi valittiin keilojen puoliarvoleveys, koska



Kuva 6.9: Lähikuva kohinataustoista A ja A' ennen ja jälkeen suodatuksen

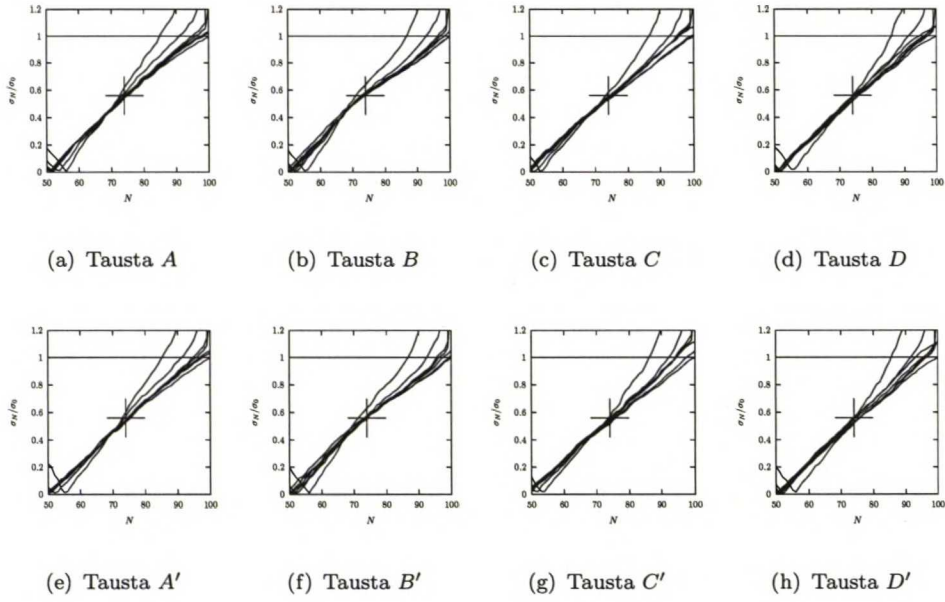
tässä vaiheessa ei ollut parempaa arvausta optimaalisesta leveydestä. Myös jokainen kohinatausta itsessään suodatettiin tällä suotimella, jotta renkaiden todellinen kohinataso voitiin laskea vertailua varten. Yhteensä testirenkaita oli siis 56. Kävi ilmi, että suodatuksen jälkeen A ja A' näyttivät melko samanlaisilta, kuten näkyy kuvasta 6.9. Sinänsä tämä ei ole yllätys, sillä meksikolaishattusuodin poistaa korkeat taajuudet kuten liukuva keskiarvoistuskinkin.

Koska huippunäytteiden määrä suhteessa muihin näytteisiin on merkityksellinen tämän testin kannalta, valittiin renkaiden pituudeksi 1500 näytettä, jolloin tulokset ovat hieman pessimistisiä.

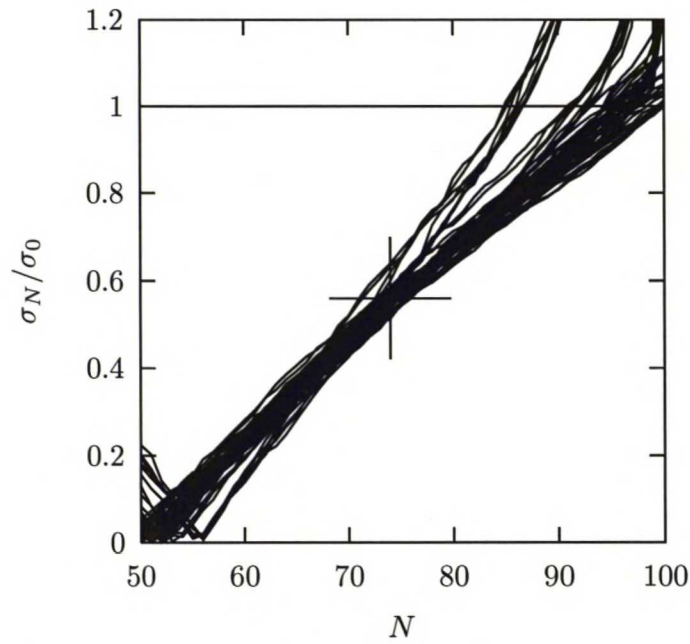
6.5.2 Testin tulokset

Jokaiselle renkaalle laskettiin σ_N kaikille kokonaislukurajoitustasoille $N \in [50, 100]$ ja tulokset jaettiin todellisella kohinatasolla σ_0 , joka on siis sama kuin pelkän kohinataustan arvo σ_{100} . Nämä suhteelliset kohinatasot on piirretty erikseen jokaiselle kohinataustalle kuvassa 6.10 ja kaikki yhdessä kuvassa 6.11.

Käyrien päämuoto on lähes lineaarinen. Meksikolaishattusuodin vaikuttaa lineaarisuuden määrään, sillä sama käyrä suodattamattomalle valkoiselle kohinalle (kuvassa 6.12) on selvästi kaarevampi. Jyrkimmin nousevat käyrät ovat 20:n kirkkaan lähteen kuvion (kuvio 2) tuottamia. Nämä myöskin eroavat eniten muista lähellä arvoa $N = 50$. Toiseksi jyrkimmin nousevat käyrät ovat

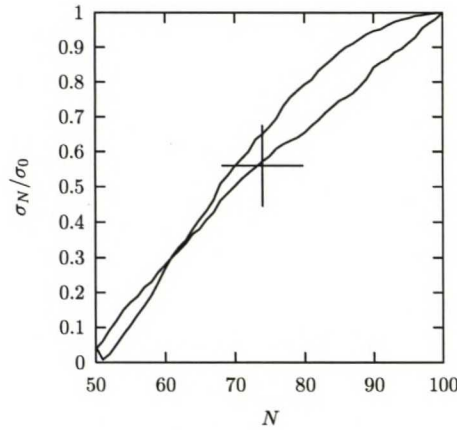


Kuva 6.10: Suhteelliset kohinasot σ_N/σ_0 eri rajoitustasolle N piirrettynä jokaiselle kohinataustalle. Jokainen käyrä esittää yhden huippukuvion tuloksia.



Kuva 6.11: Kaikki kuvan 6.10 käyrät piirrettynä yhteen kuvaan

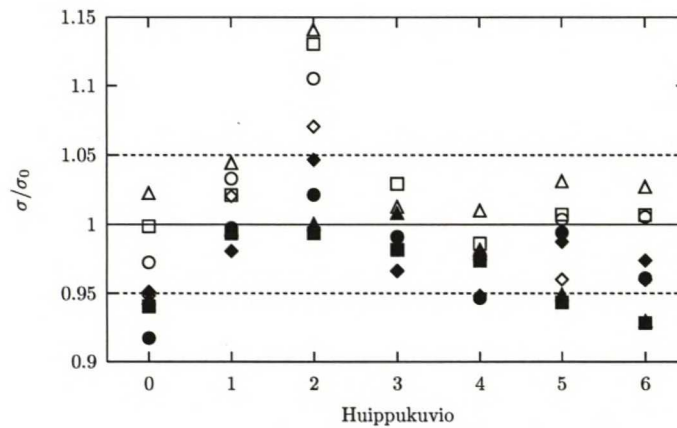
peräisin testikuvioista (kuvio 1). Näyttää siis siltä, että pahimmat tapaukset tämän menetelmän kannalta ovat ne, joissa on joko paljon ja/tai hyvin kirk-



Kuva 6.12: Suhteelliset kohinatasot suodatetulle ja suodattamattomalle taustalle A . Suodatettu kulkee lähempää ristiä.

kaita lähteitä, mikä tietysti oli odotettavissa.

Mitä suurempi N sitä parempi tarkkuus, koska menetelmän tarkkuus degeneroituu N :n lähestyessä arvoa 50 kertoimen C lähestyessä samalla ääretöntä. Toisaalta noin kohdassa $N = 75$ käyrät alkavat levitä enemmän. Optimaaliset arvot parametreille N ja $C = \sigma_0/\sigma_N$ ovat siis silmämääräisesti jossain näillä paikkeilla. Jokaiselle renkaalle laskettiin sen suhteellinen tarkkuus suorakai-teessa $N = 69; 69, 5; \dots; 78, 5$ ja $\sigma_N/\sigma_0 = 0, 48; 0, 49; \dots; 0, 64$. Optimaaliseksi pisteeksi valittiin se, jolla eri renkaiden suhteellisen virheen keskihajonta on pienin. Tämä piste on $N = 74$, $\sigma_N/\sigma_0 = 0, 56$, joka on merkitty ristillä kuviin 6.10, 6.11 ja 6.12. Tämä antaa kohinan arviointiin relaation $\sigma_0 \approx 1.786\sigma_{74}$.



Kuva 6.13: Suhteellinen tarkkuus σ/σ_0 jokaiselle testirenkaalle käytettäessä arviointia $\sigma = 1.786\sigma_{74}$. Erilaiset kohinataustat on merkitty omilla merkeillään, ja sijainti vaakasuunnassa kertoo huippukuvion; nollan kohdalla pelkkä kohinatausta.

Kaavasta $\sigma = 1.786\sigma_{74}$ saadut kohinatasot suhteessa todelliseen kohinaan (σ/σ_0) on merkitty kuvaan 6.13. Tästä nähdään, että menetelmän tuottama virhe on suurimmaksi osaksi alle 5% ja vähemmän kuin 15% kaikille käytetyille testirenkaille.

6.5.3 Johtopäätökset

Näyttää siltä, että mikäli signaalidatan ominaisuudet ovat samat kuin suodatetulla Planckin signaalidatalla, hyvä kaava kohinatason arviointiin on $\sigma = 1.786\sigma_{74}$, toisin sanoen asettamalla $N = 74$ ja $C = 1.786$ yhtälössä 6.9. Tämä menetelmä saattaa toimia muunkinlaiselle datalle, mutta silloin arvot N ja C täytyy määrittää uudelleen sillä kuvan 6.11 käyrät saattavat olla erilaiset.

Katsomalla kuvaa 6.11 huomataan myös, että lukuunottamatta kaikkein pahimpia tapauksia arvot $N = 95$ ja $C = 1$ tuottavat alle 10%:n virheen. On kuitenkin parempi käyttää yllä mainittua kaavaa, sillä se toimii myös näissä pahimmissa tapauksissa.

6.6 Huippujen etsiminen

Huipuksi katsotaan sellainen näyte, jonka arvo on suurempi kuin sitä edeltävän ja sen jälkeen tulevan näytteen arvo. Mikäli useammalla peräkkäisellä näytteellä on sama arvo — mikä tosin on *erittäin* epätodennäköistä — huipun paikaksi katsotaan näiden näytteiden paikkojen keskiarvo eli huippu on keskellä “tasankoa”.

6.7 Huippujen paikallistaminen

Etsittäessä huippuja useammista peräkkäisistä renkaista yksi huippu näkyy yleensä eri vahvuisena useassa peräkkäisessä renkaassa. Ohjelman sisällä ne muodostavat kuitenkin jokainen erillisen objektin, koska QDS käsittelee dataa rengas kerrallaan.

Mitkä näistä objekteista kuuluvat samaan lähteeseen määritellään hyvin yksinkertaisesti vain katsomalla mitkä ovat määrätyn säteen etäisyydellä toisistaan. Ryhmän kirkkain objekti määrittää lähteen paikan, ja sen koordinaatteja käytetään vertailukohtana lisättäessä uusia objekteja. Tämä saattaa aiheuttaa ongelmia, varsinkin jos käytetty säde on kovin pieni. Kuljettaessa huipun reunaa “ylämäkeen” vertailukoordinaatit siirtyvät aina uuden lisäyksen myötä, mutta ne pysyvät paikallaan kuljettaessa “alamäkeen”. Tästä johtuen viimeiset objektit saatetaan tulkita erilliseksi lähteeksi vaikka ne eivät ole sen kauempana kirkkaimmasta kuin ensimmäiset käsitellyt objektit. Lähteen paikan epätarkkuudeksi tulee näytteenleveys tai vierekkäisten renkaiden etäisyys maksimin kohdalla, riippuen siitä kumpi on suurempi.

Kun nämä objektit on kerätty yhdeksi lähdeobjektiksi järjestetään ne rengasnumeron perusteella. Kaikki sellaiset pisteet poistetaan, jotka eivät kuulu vähintään tietyn, *huippujatkuvuutena* määriteltävissä olevan määrän peräkkäisiä renkaita muodostamaan sarjaan. Jos tällaisia sarjoja on useampia, lähdeobjektille lisätään merkintä "sivuhuipuista", ja kaikki sarjat paitsi maksimivuon sisältävä sarja poistetaan. Sivuhuippuja sisältävä lähde on luultavasti seuraus siitä, että useampi lähde osuu läheisyyssäteen sisälle, tosin sivuhuiput voivat aiheutua myös taustakohinasta. Mikäli käytetään dataa, jonka renkaat eivät ole peräkkäisiä Planckin tapaan, huippujatkuvuusvaatimus tulisi poistaa käytöstä.

Mikäli maksimivuon kohta on liian lähellä järjestetyn ryhmän jompaa kumpaa ääripäätä, lähteelle lisätään merkintä siitä, että se on joko nouseva tai laskeva huippu normaalin huipun sijasta. Tällaiset huiput yleensä ylikirjoittuvat seuraavalla ohjelman ajokerralla, eikä niitä yleensä käytetä mihinkään.

6.7.1 Merkityksellinen rengasväli

Nousevat ja laskevat huiput voivat olla seurausta joko taustasignaalista tai normaaleista kohteista, jotka ovat juuri ensimmäisessä tai viimeisessä käsiteltävässä renkaassa tai hyvin lähellä niitä. Niiden esiintymistä voidaan vähentää laajentamalla käsiteltävää renkaiden väliä, mutta jatkokäsittelmällä ainoastaan alkuperäiselle käsittelyvälille osuvat löydettyt lähteet. Tällä tavalla ääripäissä sijaitsevat lähteet paikallistetaan alkuperäisen välin ulkopuolelle. Sama saattaa tapahtua myös lähteille, jotka ovat muodoltaan enemmänkin viiruja kuin pisteitä.

Jatkokäsiteltävien renkaiden väli on se, joka alunperin annetaan QDS:lle komentoriviparametrina. Sitä voidaan kutsua myös *merkitykselliseksi väliksi*, koska ainoastaan sen sisälle osuvat lähteet jatkokäsittelään ja ovat täten merkityksellisiä. QDS:lle voidaan määritellä laajennussäde, jonka verran kohdeväliä laajennetaan kummastakin päästä.

6.8 Lähteiden tunnistaminen

Yhden vastaanotinryhmän löytämät huiput kerätään yhdeksi lähdeobjektiksi edellisessä kappaleessa kuvatulla tavalla. Samaan tapaan lähdeobjektit eri vastaanotinryhmistä ryhmitellään siten, että tietyn säteen sisällä olevat muodostavat ryhmän, jonka koordinaatit ovat keskiarvo ryhmän jäsenten koordinaateista. Paikan epätarkkuus on tässä tapauksessa se säde, jonka sisällä kaikki jäsenet ovat keskiarvosta, lisättynä jäsenten paikkaepätarkkuuden maksimilla.

Seuraavaksi tätä muodostettua ryhmää verrataan QDS:n lähdetietokantaan koordinaattien perusteella, jotta saadaan selville vastaako se mitään tietokannassa olevaa lähdetä. Tämä tapahtuu tarkistamalla ovatko ryhmän koordi-

naatit tarpeeksi lähellä tietokannassa olevan lähteen koordinaatteja. Mikäli vastaavaa lähdettä ei löydy, ryhmän katsotaan kuvaavan uutta, aikaisemmin tuntematonta lähdettä. Vertailuetaisyys on sama kuin ryhmittelyssäkin.

Vertailun jälkeen lähdeobjektin tiedot tallennetaan tietokantaan osana kyseisen lähteen havaintohistoriaa. Mikäli vastaavaa lähdettä ei löytynyt, tehdään tietokantaan uusi lähde. Interaktiivisessa käyttötilassa nämä lisäykset tietokantaan pitää tehdä eksplisiittisesti, mikä mahdollistaa helpon kokeilun eri parametreilla.

Ennen lisäystä havaintohistoriaan uutta havaintoa verrataan aikaisempiin havaintoihin, jotka on tehnyt sama havainnoitsija samalla taaajuudella. Mikäli uuden ja olemassa olevan havainnon aikaleiman ero on alle tietyn ajan, tyyppillisesti luokkaa päivän, havaintojen katsotaan olevan samoja. Tämä on todennäköisesti seuraus QDS:n ajoista päällekkäin menevillä kohdeväleillä. Tällaisessa tapauksessa uutta havaintoa ei välttämättä lisätä, vaan se korvaa jo tietokannassa olevan havainnon seuraavissa tapauksissa:

1. Normaali havainto korvaa aina nousevan tai laskevan havainnon.
2. Nouseva tai laskeva eli sivuhavainto korvaa toisen sivuhavainnon mikäli sen vuo on suurempi.
3. Mikäli havainnot ovat yhtä suuret virherajojen puitteissa ($S_1 \in [S_2 - \Delta S_2, S_2 + \Delta S_2] \vee S_2 \in [S_1 - \Delta S_1, S_1 + \Delta S_1]$), havainto jolla on pienempi virhe korvaa tietokannassa olevan.

6.8.1 Osuma useampaan lähteeseen

Mikäli useampi lähde tietokannassa sopii havainnon koordinaatteihin, havainto merkitään monikäsitteiseksi ja tallennetaan tietokannassa *jokaisen* sopivan lähteen historiaan. Hälytys tehdään, mikäli *yksikin* lähteistä tuottaisi kyseisen hälytyksen.

Tulee huomata, että mikäli vertailuetaisyys on tarpeeksi pieni verrattuna keilan puoliarvovevyyteen, useaan lähteeseen sopivat havainnot ovat luultavasti enemmän tai vähemmän näiden lähteiden yhteisvaikutusta ja vuoarvot saatavat siis olla täysin käyttökelvottomia. Operaattorille tehtävä hälytys on kuitenkin hyödyllinen, sillä jotain mielenkiintoista saattaa silti olla tapahtumassa ja jälkiseuranta voidaan tehdä paremmalla erottelukyvällä.

6.9 Planeettojen käsittely

Jotkut Maan kiertoradan ulkopuolella olevat aurinkokunnan kappaleet (planeetat, kuut, asteroidit; tässä näitä kaikkia kutsutaan planeetoiksi) säteilevät

radioalueella ja ne ovat Planckin keilanleveyksillä pistemäisiä⁸. Ainakin osan säteilyvuo on muuttumaton ja hyvin tunnettu, joten niitä voidaan käyttää instrumenttien kalibrointiin mikäli niiden etäisyys satelliitista tiedetään tarpeeksi tarkasti.

Planeettojen käsittely on ongelmallista, sillä toisin kuin ekstragalaktiset tai galaktiset pistelähteet, ne liikkuvat havaittavasti Planckiin nähden ja niiden etäisyyden ja sijainnin määrittämisessä tarvitaan huomattavan tarkkaa tietoa sekä niiden että Planckin kiertoradasta [Ma2006].

Koska planeettojen huomioonottaminen täysin itsenäisesti QDS:n sisällä olisi huomattavan monimutkaista, näin ei tehdä vaan planeettojen käsittelyyn käytetään yhtä tai useampaa seuraavista menettelytavoista:

1. Jätetään planeetat käsittelemättä kokonaan, jolloin planeetat näkyvät normaalina pistelähdehavaintona. Tällöin jää ohjelmiston operaattorin vastuulle päätellä muita apukeinoja käyttäen mitkä näistä ovat planeettoja.
2. Yksinkertainen arvaus kirkkauden perusteella. Jotkut planeetat ovat hyvin kirkkaita, jopa tuhansia Janskyja, joten vuolle voidaan asettaa tietty kynnysarvo jota kirkkaammat lähteet luokitellaan planeetoiksi. Tämä ei kuitenkaan auta niiden planeettojen kohdalla, joiden kirkkaus ei ylitä kynnysarvoa mutta jotka kuitenkin erottuvat kohinasta.
3. Planeettojen näkyminen tietyssä vastaanottimissa tullaan todennäköisesti datankäsittelykeskuksessa merkitsemään vastaanotindataan asiainkuuluvien näytteen osalta. Näiden merkintöjen perusteella kyseiset näytteet voidaan joko jättää pois tai niiden perusteella pistelähde voidaan luokitella planeetaksi ja tehdä kalibraatioehdotus.
4. Käytetään ulkoista tietokantaa tai ohjelmaa, jolle voidaan tiedusteluna lähettää aikaleima ja suuntaus ja joka palauttaa vastaukseksi onko kyseisessä kohdassa planeetta ja jos on, kuinka kirkas sen pitäisi olla.

Mikäli mahdollista, QDS tulee käyttämään jompaakumpaa kahdesta jälkimmäisestä vaihtoehdosta. Kumpaan voidaan käyttää pistelähteen tunnistamiseen planeetaksi. Kalibrointikäytössä ongelmaksi saattaa muodostua se, että keskiarvoistettujen datarenkaiden aikatarkkuus on yksi tunti, jonka aikana planeetta ei välttämättä ollut koko ajan vastaanottimen keilassa ja renkaassa näkyvä kirkkaus on näin ollen heikompi kuin planeetan todellinen kirkkaus. Tätä vielä monimutkaistaa se, että samassa kohtaa saattaa olla joku muukin pistelähde.

⁸Mars alle 0,4 kaariminuuttia, Jupiter alle 0,8 kaariminuuttia

Luku 7

Testaus

Ohjelmisto oli edellisen kehittäjän jäljiltä todettu testaamalla ainakin pääosin toimivaksi. Sille on kehityksen mukana tehty jatkuvaa toimivuustestausta, eli aina muutosten tai lisäysten jälkeen on tarkistettu simulaation tuottamaa testidataa käyttäen, että ohjelmisto toimii niin kuin pitäisikin. Käytännössä tämä tarkoittaa sitä, että simulaattorilla on generoitu dataa tietyillä pistelähteillä ja taustalla, ja tätä dataa on käytetty QDS:n testidatana jotta nähdään, havaitseeko se pistelähteet oikean suuruisena.

Ohjelmiston tälle versiolle ei ole kuitenkaan vielä tehty *suorituskykytestausta*, jossa testattaisiin kuinka tehokkaasti se toimii ja *kuinka hyvin* se suoriutuu päätehtävästään eli pistelähteiden erottelusta. Näitä testejä varten simulaattorilla generoidaan koko taivaan data käyttäen erilaisia pistelähdekokoelmia. Tarkoitus olisi käyttää ainakin testikokoelmaa ja WMAP:iin ja muihin havaintoihin pohjautuvaa oikeita pistelähteitä sisältävää kokoelmaa sekä mahdollisesti myös täysin satunnaisesti generoitua kokoelmaa. Nämä annetaan syötteeksi QDS:lle ja tuloksista voidaan erotella ainakin seuraavat seikat:

- Valheellisten havaintojen lukumäärä ja sen riippuvuus havaintokynnyksestä.
- Havaitsematta jääneiden kohteiden lukumäärä ja sen riippuvuus havaintokynnyksestä ja kohteen kirkkaudesta.
- Kuinka hyvin havaitun kohteen kirkkaus vastaa kokoelmassa annettua oikeaa kirkkautta.
- Huomattiinko hälytyskriteerit täyttävät kohteet kuten esimerkiksi lähteet, joilla on käänteinen spektri.

7.1 Planck-simulaattori

QDS:n testaukseen on käytetty Planck-simulaattorin tuottamaa dataa. Simulaattori on osa Planckin datankäsittelykokonaisuutta ja sen tarkoitus on mal-

lintaa Planckin tuottama data mahdollisimman hyvin, jotta sitä voidaan käyttää datan analyysimetodien sekä datankäsittelykeskusten infrastruktuurin testaamiseen.

Simulaattori mallintaa erikseen jokaisen mikroaaltotaivaan komponentin — kuten taustasäteily, pistelähteet ja kohina — ja yhdistää ne lopulta havaintodataksi. Simulaatioon voidaan ottaa mukaan vain halutut komponentit, ja kunkin komponentit parametrit ovat erikseen säädettävissä. Osa komponenteista luodaan kaksiulotteisina taivaskarttoina, jotka yhdistetään yhdeksi taivaskartaksi, jota simulaattori kuvaa käyttäen Planckin kuvausstrategiaa. Jotkin komponentit, kuten kohina, lisätään vasta tässä kuvausvaiheessa. Generoitu data on aikajärjestettyä dataa renkaina, mutta simulaattoriohjelmisto sisältää myös mahdollisuuden muuntaa se taivaskartoiksi. Simulaattorissa voi valita, generoidaanko jokaista rengasta kohden kaikki 60 kuvausrengasta vai vain yksi ideaalisesti kuvattu rengas. QDS:n kanssa käytetään ensimmäistä tapaa, jotta QDS:lle syötetty data olisi mahdollisimman realistista ja jotta ohjelman kaikki osat voidaan testata. Simulaattorin eri moduulit on kuvattu sen käyttöohjeessa [Re2006].

Simulaattori on lähdekoodimuodossa jaettava unix-ohjelmistopaketti, joka toimii useissa eri järjestelmissä. Tärkeimmät vaatimukset simulaattorin käyttämiselle asettavat sen kohtuullisen suuri muistin käyttö sekä sekä generoidun datan vaatima huomattavan suuri levytila. Koska kyseessä on laskentaintensiivinen ohjelma, on myös riittävä suoritinteho suotavaa. Tätä diplomityötä tehtäessä simulaattoria on ajettu Metsähovin Linux-PC -koneissa.

Luku 8

Yhteenveto

Tässä diplomityössä toteutettiin Planck-satelliitin datasta pistelähteitä erotteleva ohjelmisto. Sen tarkoituksena on käsitellä Planckin matalataajuusinstrumentin dataa sitä mukaa kun se lähetetään maahan ja erotella siitä mielenkiintoiset ekstragalaktiset pistelähteet. Tällainen lähes reaaliaikainen datankäsittely on tarpeen, koska aktiivisten galaksien muuttuvuuden aikaskaala voi olla jopa päiviä ja Planckin varsinaiset tulokset saadaan vasta useita kuukausia koko taivaan kuvauksen jälkeen. Ohjelmisto on tarkoitettu pääasiassa Metsähovin radiotutkimusaseman käyttöön ja sitä tullaan ajamaan Planckin LFI-datankäsittelykeskuksessa. Planck laukaistaan alkuvuodesta 2008, minkä jälkeen on matka kiertoradalle L2-pisteeseen ja testivaihe. Varsinainen taivaan kartoitus alkaa noin puoli vuotta laukaisun jälkeen, jolloin myös QDS aloittaa koko satelliitin toiminta-ajan kestävän toimintansa.

8.1 Tavoitteiden täyttyminen

Toimivuustestauksen perusteella ohjelmisto on todettu toimivaksi: se löytää pistelähteet signaalista. Tavoitteissa esitetyllä 30%:n tarkkuudella havaintokynnys tulee simulaattorin perusteella olemaan hieman yhden Janskyn alapuolella. Havaittujen pistelähteiden paikkatarkkuus on käyttötarkoitukseensa riittävä. Yhden vuorokauden datan käsittely QDS:llä kestää vain joitakin minuutteja, joten ainoa aikarajaan vaikuttava tekijä on Planckin LFI-datankäsittelykeskus, jonka senkin pitäisi tämän hetken tietojen mukaan toimia tarpeeksi nopeasti QDS:lle asetetun viikon aikarajan puitteissa.

Hälytyskriteerit ovat kaikki hyvin määriteltyjä epäyhtälöitä, joiden toimivuus on helppo todeta yksinkertaisin testein. Hälytykset toimivat mikäli QDS:n muut osat toimivat, ainoaksi mahdolliseksi ongelmaksi jää hyvien arvojen määrittäminen kriteereissä esiintyvillä parametreille. Toisaalta edes löysästi valituilla parametreilla hälytysten määrä vuorokaudessa ei tule olemaan kohuttoman suuri.

Periaatteessa QDS:n perimmäinen tavoite on erotella pistelähteet taustas-

ta “mahdollisimman hyvin”. Tämän toteutuminen tarkoittaisi käytännössä viimeiseen hetkeen jatkuvaa tutkimusprosessia, joten tässä vaiheessa siitä ei voi sanoa mitään. Voitaneen kuitenkin todeta, että QDS suoriutuu tehtävästään *tarpeeksi* hyvin.

8.2 Tulevaisuus

Tulevaisuudensuunnitelmiin itse ohjelmiston kannalta kuuluvat laajamittainen testaus, uusien ominaisuuksien lisääminen sekä tuki Planckin LFI-datankäsittelykeskuksen rengastietokannalle. Koska ohjelma on jo todettu teknisesti toimivaksi, testauksella pyritään pääasiassa löytämään optimaaliset arvot QDS:n konfiguraatioparametreille.

Uudet ominaisuudet liittyvät pääasiassa suodatukseen. Suunnitelmissa on toteuttaa ainakin meksikolaishattuaallockeen optimaalisen leveyden automaattinen määrittäminen sekä tuki optimoidulle suotimelle ja toisen (ja mahdollisesti korkeampienkin) asteen meksikolaishattusuotimelle. Näiden ominaisuuksien todellinen hyöty selviää vasta testauksen jälkeen.

Planeettojen käsittely olisi tarkoitus toteuttaa jollain kappaleessa 6.9 esitetyllä tavalla. Tällä hetkellä planeettoja ei käsitellä, vaan ne näkyvät uusina lähteinä ja niiden tunnistus on operaattorin vastuulla.

Myös ohjelman tällä hetkellä hieman kömpelö interaktiivinen käyttöliittymä kaipaa parannusta. Vaihtoehtoina on joko sen parantaminen tai WWW-pohjaisen käyttöliittymän toteuttaminen sen tilalle tai rinnalle. Tässä päädytään luultavasti jälkimmäiseen vaihtoehtoon WWW-käyttöliittymän sisältämien etujen takia.

Kirjallisuutta

- [Ba2003] Barreiro, R. B., Sanz, J. L., Herranz, D., Martínez-González, E.: *Comparing filters for the detection of point sources*. Monthly Notes of Royal Astronomical Society 342:119–133, 2003.
- [Ba2004] Barreiro, R. B., Hobson, M. P., Banday, A. J., Lasenby, A. N., Stolyarov, V., Vielva, P., Górski, K. M.: *Foreground separation using a flexible maximum-entropy algorithm: an application to COBE data*. Monthly Notes of Royal Astronomical Society 351:515–540, 2004.
- [Bluebook] The Planck Consortia:
Planck: The Scientific Programme (the “Bluebook”)
ESA-SCI(2005)1, ESA Publications 2005.
(Saataavilla Planck-projektin WWW-sivuilta)
- [Fa1993] Fayyad, U. M., Weir, N., Djorgovski, S.: *Automated cataloging and analysis of sky survey image databases: the SKICAT system*. Proceedings of the second international conference on Information and knowledge management, ACM Press 1993.
- [Gó2005] Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., Bartelmann, M.: *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*. The Astrophysical Journal 622:759–771, 2005.
- [Go2006] González-Nuevo, J., Argüeso, F., López-Caniego, M., Toffolatti, L., Sanz, J. L., Vielva, P., Herranz, D.: *The Mexican hat wavelet family: application to point-source detection on cosmic microwave background maps*. Monthly Notes of Royal Astronomical Society 369:1603–1610, 2006.

- [Ha2001] Hanisch, R. J. et al.:
Definition of the Flexible Image Transport System (FITS).
Astronomy & Astrophysics 376:359–380, 2001.
<http://fits.gsfc.nasa.gov/>
- [He2002a] Herranz, D., Gallegos, J., Sanz, J. L., Martínez-González, E.:
Point source detection and extraction from simulated Planck time-ordered data using optimal adaptive filters.
Monthly Notes of Royal Astronomical Society 334:533–541, 2002.
- [He2002b] Herranz, D., Sanz, J. L., Barreiro, R. B., Martínez-González, E.:
Scale-adaptive Filters for the Detection/Separation of Compact Sources
The Astrophysical Journal 580:610–625, 2002.
- [Ja1996] Janssen, M. A. et al.:
Direct Imaging of the CMB from Space.
[astro-ph/9602009], 1996.
- [Ló2005] López-Caniego, M., Herranz, D., Barreiro, R. B., Sanz, J. L.:
Filter design for the detection of compact sources based on the Neyman-Pearson detector
Monthly Notes of Royal Astronomical Society 359:993–1006, 2005.
- [Ló2006] López-Caniego, M., Herranz, D., González-Nuevo, J., Sanz, J. L., Barreiro, R. B., Vielva, P., Argüeso, F., Toffolatti, L.:
Comparison of filters for the detection of point sources in Planck simulations
Monthly Notes of Royal Astronomical Society 370:2047–2063, 2006.
- [Lä1999] Lähteenmäki, A., Valtaoja, E.:
Total Flux Density Variations in Extragalactic Radio Sources. III. Doppler Boosting Factors, Lorentz Factors, and Viewing Angles for Active Galactic Nuclei.
The Astrophysical Journal 521:493–501, 1999.
- [Ma2002] Maino, D., Burigana, C., Górski, K. M., Mandolesi, N., Bersanelli, M.:
Removing $1/f$ noise stripes in cosmic microwave background anisotropy observations.
Astronomy & Astrophysics 387:356–365, 2002.
- [Ma2006] Maris, M., Burigana, C.:
Planck LFI — On Moving Objects Solar System Calculations
Planck internal document PL-LFI-OAT-TN-034, 2006.

- [Pa2004] Parviainen, Mikko:
Pistelähteiden nopea erottelu Planck-satelliitin datasta.
Diplomityö, Teknillinen korkeakoulu, 2004.
- [Planck] Planck-projektin WWW-sivut
<http://www.rssd.esa.int/Planck>
- [Re2006] Reinecke, M., Dolag, K.:
Compilation and Usage of the Planck Simulation Modules
Planck internal document PL-COM-MPA-MA-SIM007, 2006.
- [Sa2001] Sanz, J. L., Herranz, D., Martínez-González, E.:
Optimal Detection of Sources on a Homogenous and Isotropic Background
The Astrophysical Journal 552:484–492, 2001.
- [Sz2000] Szalay, A. S., Kunszt, P. Z., Thakar, A., Gray, J., Slutz, D., Brunner, R. J.:
Designing and mining multi-terabyte astronomy archives: the Sloan Digital Sky Survey
Proceedings of the 2000 ACM SIGMOD international conference on Management of data, ACM Press 2000.
- [Sz2002] Szalay, A. S., Gray, J., Thakar, A. R., Kunszt, P. Z., Malik, T., Raddick, J., Stoughton, C., vandenBerg, J.:
The SDSS SkyServer—Public Access to the Sloan Digital Sky Server Data
Proceedings of the 2002 ACM SIGMOD international conference on Management of data, ACM Press 2002.
- [Vi2001] Vielva, P., Martínez-González, E., Cayón, L., Diego, J. M., Sanz, J. L., Toffolatti, L.:
Predicted Planck extragalactic point-source catalogue.
Monthly Notes of Royal Astronomical Society 326:181–191, 2001.
- [Vi2003] Vielva, P., Martínez-González, E., Gallegos, J. E., Toffolatti, L., Sanz, J. L.:
Point source detection using the Spherical Mexican Hat Wavelet on simulated all-sky Planck maps
Monthly Notes of Royal Astronomical Society 344:89–104, 2003.